

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:53:45 : Search time 4457.97 seconds
(without alignments)
3382.078 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggTTtaattaccCaagtttg.....taaatacgattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: gb_est1.*
21: gb_est2.*
22: gb_est3.*
23: gb_est4.*
24: gb_est5.*
25: gb_est6.*
26: gb_est7.*
27: gb_est8.*
28: gb_est9.*
29: em_est10.*
30: gb_est11.*
31: gb_est12.*
32: gb_est13.*
33: gb_est14.*
34: gb_est15.*
35: gb_est16.*
36: gb_est17.*
37: gb_est18.*
38: gb_est19.*
39: gb_est20.*
40: gb_est21.*
41: gb_est22.*
42: gb_est23.*
43: gb_est24.*
44: gb_est25.*
45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: em_est20.*
53: em_est21.*
54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*
59: gb_est33.*
60: gb_est34.*
61: gb_est35.*
62: gb_est36.*
63: gb_est37.*
64: gb_est38.*
65: em_est27.*
66: em_est28.*
67: em_est29.*
68: em_est30.*
69: gb_est39.*
70: gb_est40.*
71: gb_est41.*
72: gb_est42.*
73: gb_est43.*
74: gb_est44.*
75: em_est31.*
76: em_est32.*
77: em_est33.*
78: em_est34.*
79: gb_est45.*
80: gb_est46.*
81: gb_est47.*
82: em_est35.*
83: em_est36.*
84: em_est37.*
85: gb_est48.*
86: gb_est49.*
87: gb_est50.*
88: gb_est51.*
89: gb_est52.*
90: gb_est53.*
91: gb_est54.*
92: gb_est55.*
93: gb_gss1.*
94: gb_gss2.*
95: gb_gss3.*
96: gb_gss4.*
97: em_gss1.*
98: em_gss2.*
99: em_gss3.*
100: em_gss4.*
101: gb_gss5.*
102: gb_gss6.*
103: gb_gss7.*
104: gb_gss8.*
105: gb_gss9.*
106: em_gss5.*
107: em_gss6.*
108: em_gss7.*
109: em_gss8.*
110: em_gss9.*
111: em_gss10.*
112: em_gss11.*
113: gb_gss10.*
114: gb_gss11.*
115: em_gss12.*
116: gb_gss12.*

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.

TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3036136.

CONTACT: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone="yk15h12"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 123 a 63 c 100 g 73 t 1 others
ORIGIN

Query Match 10.5%; Score 359; DB 81; Length 360;
Best Local Similarity 99.7%; Pred. No. 4.2e-95;
Matches 359; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 773 atcgattgaagatgtccagagacagaaaccaactcatttgggtgaacacaaagtccaaa 832
|||||
Db 1 ATCCATTGAAGATGCTCCAGAGACGANACCAACTCAITTTGGTTGACCAAAAGTCCAAA 60

QY 833 aagcccgagaatgaaagaggagcttctcggttactagttaacatgtctccgagtga 892
|||||
Db 61 AAGCCCGCAGAATATGAAGAGGAGCTTCTTCGGTTACTAGTTAACATGTCTCCGAGTGA 120

QY 893 agttgaacggttaaaagataaaaatcaggagcatgttcagcgacgaatggccatcag 952
|||||
Db 121 AGTTCAACGGTTAAAGAAATAAAAAATCAGGAGCAIGTTACGCGACGAATGGCCATCGAG 180

QY 953 gagtaaggagaagcgcggaagattgtgattcaggagacagcggaagggatgaagatga 1012
|||||
Db 181 GAGTAAGGAGAAGCGCGGAAGATTGTGATTTCAGGAGACACGCGAAGGGATGAAGATGA 240

QY 1013 ggatgatgagagatgtattcccggggagactatgtctcagggaactactatttgtcg 1072
|||||
Db 241 GGATGATGAGGATAGTATTCCCGGGGAGACTATGTCTCAGGGAACTACTATTATTGTTCG 300

QY 1073 aaagacaaaacagcgctgtacggcacacaattctcatcgaaaagaatagatctc 1132
|||||
Db 301 AAGACAAAACCCAGCGCTGTACGGCACAACTCTCATGAAAAGAAGTATAGATGCTC 360

RESULT 3
LOCUS D68853 360 bp mRNA EST 07-DEC-1995
DEFINITION CELK056DZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk56d12 5', mRNA sequence.

ACCESSION D68853
VERSION D68853.1 GI:1104494
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Eukaryota; Metazoa; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 378)
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.

TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)

REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.

TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188359.

CONTACT: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone="yk56d12"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 97 a 72 c 96 g 90 t 5 others
ORIGIN

Query Match 10.4%; Score 355; DB 81; Length 360;
Best Local Similarity 98.6%; Pred. No. 6.4e-94;
Matches 355; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2540 cctaataagacaatgacggtctcgatgatgaagtacgagctgattgtttcaaacatttt 2599
|||||
Db 1 CCTAAATGACAATGACCGTCTCGATGATGAAGTAGACGTGGTTGATGTTTCAAGACITTT 60

QY 2600 ggtgacaatttcaacgcagtgctgccttcttgaactaatgagaaggatgagtcagcga 2659
|||||
Db 61 GGTCACAATTTCAACGCAGTGCCTGCGTCATTTTGACTAATCAGAAGGATGAGTCAGCGAA 120

QY 2660 attcgaaccttgatctctcgaacgagatgctgttgatgacatggtgacgacgttct 2719
|||||
Db 121 ATTCCGAACCTTGGATCTCTCGAAACGGAGATGCTTGTGTGCACATGGTGAGCGACGTTCT 180

QY 2720 gacatgtggaatctattggaggagtaacaadaalagactgcacgcgaacactattcact 2779
|||||
Db 181 GACATGTGGATCTATTGGAGGAGTAAACAAATGAGCTGGCAGCGCAACTATTCTACT 240

QY 2780 gatcaggaactgtccgcgaagattttgacagacaactctgggtttggcggttggccacgc 2839
|||||
Db 241 GATCAGGAACCTGTCCNCCGAAGATTNTGACACAGACAACTCTGGGTTTGGCGGTGGCCACGC 300

QY 2840 gttgtgtcgcgaagatttgcatagatgaccgagattccccgaaagtcagtcacaaacgt 2899
|||||
Db 301 GTTGTGTCTGCGAAGATTTGNATAGATGACCGGANATTCCTCCGAAAGTTCAGNCAATAGCT 360

RESULT 4
LOCUS C11998 378 bp mRNA EST 28-DEC-1998
DEFINITION C11998 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk144h7 5', mRNA sequence.

ACCESSION C11998.1 GI:1559551
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 378)
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.

TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)

COMMENT On Nov 2, 1998 this sequence version replaced gi:3828838.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
Location/Qualifiers
Source
1..378
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk144h7"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
93 a 96 c 89 t 10 others

Query Match 10.3%; Score 352.8; DB 80; Length 378;
Best Local Similarity 96.6%; Pred. No. 2.9e-93;
Matches 365; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
Qy 2757 ctggcagcgcaacactattcactgatcaggaactgtccgcgaagatitlgacagacaat 2816
Db 1 CTGGCAGCGCACACTATTCTACTGATCAGGAAGTGTCCGCCGAAGATNTTGACACACAA 60
Qy 2817 ctgggtttggcgttggccacgcgttgtgtgtcgcgaagatttgcataagatgaccgag-a 2875
Db 61 CTGGGTTTGGCGGTGGCCACGCGTGTGTGTCGCCAAGATTGNTATAGATGACCGAGCA 120
Qy 2876 ttccccgaagtcaatcagtcgtgtgattcaacaagaagtcgctcgaatccctccg 2935
Db 121 TTCCCGAAGATCAGTCAATACGTGTGCATNCACAAAGAAAGTGCCTCGACTCCCTCCG 180
Qy 2936 actatttccacatcatcgagacacaggtgtgtgtctggaattcgaagtgacacg 2995
Db 181 NCTATTCTCCACATCATCGNGAGCATCAGGTGTGTGTCTGGAATTCAGGAAGGTACACG 240
Qy 2996 cgaatggcctacgaatgattgaactcgtcgtcgcacgcgtggttccaatctatt 3055
Db 241 CGGAATGGCCTACGAATGGATTATGANTCGCTGCTCGACGCGGTCCCAATCTATT 300
Qy 3056 cgcacgaaacctactgacacaaagcttcaaggacaaatccacgttttagtacgcttta 3115
Db 301 CGCATCGAAACCTTACTNGACACANAGCTTCAAGGGACANTCCACGTTTAGTACGCTTTA 360
Qy 3116 tcaagaggcgtataatca 3133
Db 361 TCANGAGGTGATATCA 378

RESULT 5
D67853 360 bp mRNA EST 07-DEC-1995
LOCUS CELK087G4F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk87g4 5', mRNA sequence.
ACCESSION D67853
VERSION D67853.1 GI:1105632
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuiki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3035947.

Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
Location/Qualifiers
Source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk87g4"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
98 a 86 c 85 g 88 t 3 others

Query Match 10.3%; Score 350.6; DB 81; Length 360;
Best Local Similarity 98.1%; Pred. No. 1.3e-92;
Matches 353; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1303 gattaccctaccgggaacaaattcaatacactgaatactctgccgatcacccgtcgaa 1362
Db 1 GATTACCGTACC CGGAACCAATTCAATACACTGAATACTCTGCCGATCACCCGTCGAAT 60
Qy 1363 catctccttccaccctagaaatgagagagaaaaacgatcacgaatgagacacacactccta 1422
Db 61 CATCTCCCTTCTCCACTAGAAATGAGAGAAAACGATCAGGAATGAGCACACGACTCCTTA 120
Qy 1423 tgaagaatggaaactagatgattctcgaaagttacccttttttcgattgctccttagcag 1482
Db 121 TGAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCTNTGTGGATGCTCTTAGNAG 180
Qy 1483 ttctgatttttaaccgattggattgctcgtcgtggaagtcgatattctcaaaagccgctg 1542
Db 181 TTCTGATTTTAAATCCGATTGGATTGCTGCTGGAAGTGCATATTCTCAAAGCGGNTG 240
Qy 1543 cagaagctccgattcctcccgcttcgagcatggagagatgattgaccccgatggaa 1602
Db 241 GAGAAGCTCCGATTGCCCTCCCGCTCGAGCATGGAAGAGTGTATGATGACCCGGATGGAA 300
Qy 1603 ctgacactcggacgcgttttcttggaaggagtgatcatcaatatgagctatgtctgggtgt 1662
Db 301 CTAGCACTCGGACGCTTCTCTGGGAAGGAGTATCATCAATATGAGCTATGCTGGGTGT 360

RESULT 6
D34815 360 bp mRNA EST 08-AUG-1994
LOCUS CELK012HF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk12hl2 5', mRNA sequence.
ACCESSION D34815
VERSION D34815.1 GI:526331
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuiki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3035947.

Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="yk12h12"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 119 a 106 c 58 g 74 t 3 others
ORIGIN

Query Match 10.2%; Score 347.4; DB 81; Length 360;
Best Local Similarity 97.5%; Pred. No. 1.1e-91;
Matches 351; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 545 gaattcactcatgaatgctatcttcaccacaccatcctcaatgacttctcaaccata 604
Db 1 GAATTTACTCATGATGCTATCTCCACCACACCATACCTCAATGACTCTCTCAACCATA 60
Qy 605 tacagaagccatggagcatatacaacgggtacatgtctccatcacgaccaagctcaaggccc 664
Db 61 TACGAAGCAATGGAGCATATCAACGGGTACATGTCTCATACGACCAAGCTCAAGGNCC 120
Qy 665 atcaggaccatcatattactcacaaacaccatcaatctccacacccatcaccaccatca 724
Db 121 ATCAGGACCATCATATTACTCACAAACACCATCAATCTCCACCCTCATCACCAACCATCA 180
Qy 725 ccaccgatgcgcaaaaatccatgagaacccctgaacagtgcatctccatcgattgaaga 784
Db 181 CCACCCGATGCCAAAATCCATGAGAACCCCGAACAACTGGNATCTCCATCGATTGAAGA 240
Qy 785 tgctccagagacgaaacccaactcatcttggttgaaccacaaagtccaaaagcccgagaa 844
Db 241 TGCTCCAGAGACGAAACCAACTCATTTGNTGAACACAAAAGTCCAAAAGCCCGCAGAA 300
Qy 845 tatagaagagagcttcttcggttactagttaacatgtctccgagtgaagttgaacggtt 904
Db 301 TATGAAGAGAGGCTTCTTCGGTTACTAGTTAAACATGTCTCCGAGTGAAGTTGGACGGTT 360

RESULT 7
C63905
LOCUS C63905 360 bp mRNA EST 22-SEP-1997
DEFINITION C63905 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk349h3 5', mRNA sequence.
ACCESSION C63905
VERSION C63905.1 GI:2422610
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
On Apr 7, 1998 this sequence version replaced gi:3035983.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"

/db_xref="taxon:6239"
/clone="yk349h3"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 121 a 92 c 73 g 66 t 8 others
ORIGIN

Query Match 9.9%; Score 340; DB 81; Length 360;
Best Local Similarity 97.5%; Pred. No. 1.8e-89;
Matches 351; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Qy 642 ccatacgaccaagctcaaggcccatcaggaccatcatattactcaaacaccatcaatct 701
Db 1 CCATACGACCAANCTCAAGGCCATCAGACCATCATATTACTCAACACCATCAATCT 60
Qy 702 ccaccacctcatccaccatccaccctcgatgccaaaatccatgagaacctg-aaca 760
Db 61 CCACCACCTONTACCACCATCACCCCGATGCCAAAATCCATCAGAACCTCTGAACA 120
Qy 761 agtggcatctccatgattgaagatgctccagagagacgaacaaactatttggttgaacc 820
Db 121 AGTGCATCTCCATCGATTGAAGATGCTCCAGAGACGAAACCAACTCATTTGGTTGAACC 180
Qy 821 acaaaagtcgcaaaagcccgagaaatatgaagaggagcttcttcggttactagttaaacat 880
Db 181 ACAAAGTCCAAAAGCCCGCAGATATNAAGAGAGAGCTTCTTCGGTTACTAGTTAAAGAT 240
Qy 881 gtctccgagtgaagtgaacggttaagaataaaaaatcagagcatgtctcagcgacgaaa 940
Db 241 GTCTCCGAGTGAAGTTGAACGTTAAAGAAATAAAAATCNGAGCATGTTCAGCCACGCA 300
Qy 941 tgggcatcagagagtaagaggagggcgcggaagattgtgattcaggagagacgcggaagg 1000
Db 301 TGGGCATCGAGGAGTANGGGAANGCGGGAANATTGTGATTAGGAGACAGCGGAANG 360

RESULT 8
D64806/c
LOCUS CELK012GZR Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk12g12 3', mRNA sequence.
ACCESSION D64806
VERSION D64806.1 GI:1115082
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 329)
Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
Toward an expression map of the C.elegans genome
Unpublished (1994)
On Jun 22, 1998 this sequence version replaced gi:3246664.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..329
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="yk12g12"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"

BASE COUNT 81 a 76 c 95 t 1 others
ORIGIN /dev_stage="varied"

Query Match 9.5%; Score 326; DB 81; Length 329;
Best Local Similarity 99.1%; Pred. No. 2.4e-85;
Matches 326; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3066 cactactggacacaaagcttcaaggagacaaatccacgcttttagtacgctttatcaagagcg 3125
|||||
DB 329 CCTACTGACACAAAGCTTCAAGGAGACAAATCCACGTTTAGTACGCTTTATCAAGAGCG 270
|||||
QY 3126 tataatcattatgcattatttaattggacaaaggagattgttgagactatttgcctac 3185
|||||
DB 269 TATATCATATTGCCATATTATTAATGGACAAAGGAGATTGTGGAGACTATTGTCTAC 210
|||||
QY 3186 gagctcacgtgcgaatgtctaacgagacaaacccacacacgacgtggtcagggcyccga 3245
|||||
DB 209 GAGCTCACGTGCGGAATGCTCAACGGAGCAACCCACAAAGCCACGTGGTCAGGCGTCCGA 150
|||||
QY 3246 cgcgttcgactctacaaaatgagcgcgttcgaggaagagtgagcatgcgacgctcgct 3305
|||||
DB 149 CCGGTTGCGATCTACAAATGAGACCGCGTCCGAGGAAGAGTGGCATGCGCGCT 90
|||||
QY 3306 caaccggacgcatcttcatactacacgtgttaaaactacatacttctatgatctttga 3365
|||||
DB 89 CAACCGGAGCCGCAATTCATCTTCATACACGTGTTAACTACATACACTCTATGATCTTTGA 30
|||||
QY 3366 attgaacaaaaatgattttattcagaat 3394
|||||
DB 29 ATTGAACAAATGATTTTATTTCAGAAAT 1
|||||

RESULT 9

D64735/c
LOCUS CELK087G4R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk87g4 3', mRNA sequence.
ACCESSION D64735 328 bp mRNA EST 13-DEC-1995
VERSION D64735
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 328)
AUTHORS Kohara,Y., Mitsuiki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Oct 6, 1998 this sequence version replaced gi:3705561.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: yokohara@lab.nig.ac.jp.

FEATURES

source
1..328
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk87g4"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 81 a 75 c 91 t 6 others
ORIGIN

Query Match 8.9%; Score 305.8; DB 81; Length 328;
Best Local Similarity 97.0%; Pred. No. 2.2e-79;
Matches 318; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 3068 ctactggacacaaagcttcaaggagacaaatccacgcttttagtacgctttatcaagagcgta 3127
|||||
DB 328 CTACTGGACACAAAGCTTCAAGGAGCAATCCACGTTTAGTACGCTTTATCAAGAGCGTA 269
|||||
QY 3128 taatcattatgcattatttaattggacaaaggagattgttgagactatttgcctacga 3187
|||||
DB 268 TAATCATATTGCGATTATTAAATGGCAAGGGAGATTGTTGAGACTATTGTCTNCGA 209
|||||
QY 3188 gctcagtcgccaatgtctaacgagacaaacccacacacgctgtcagggcyccgacg 3247
|||||
DB 208 GCTCAGTGNCGAATGCTCACCGGAGCCAAACCCACAAAGCCACGTGTCAGGCGTCCGACG 149
|||||
QY 3248 cgttcgactctacaaaatgagcgcgttcgaggaagagtgagcatgcgacgctcgctca 3307
|||||
DB 148 CGTTGATCTACAAATGAGCGCGTCCGAGGAGAGTGAGCATNCGACGCTCGGCTCA 89
|||||
QY 3308 accgacgcatcttcatacttcatactggttaaaactacatactt-ctatggatctttgaa 3366
|||||
DB 88 ACCGGAGCGATTTCATCTTCATACACTGTTAACTACATACACTTNTATGATCTTTGAA 29
|||||
QY 3367 ttgaacaaaaatgattttattcagaat 3394
|||||
DB 28 TTGAACNACAATGATTTTATTTCAGAAAT 1
|||||

RESULT 10

C10363/c
LOCUS C10363 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk144h7 3', mRNA sequence.
ACCESSION C10363
VERSION C10363
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 252)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Jul 8, 1999 this sequence version replaced gi:5422657.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: yokohara@lab.nig.ac.jp.

FEATURES

source
1..252
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk144h7"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 57 a 63 c 62 g 66 t 4 others
ORIGIN

Query Match 6.7%; Score 229.2; DB 80; Length 252;
Best Local Similarity 96.0%; Pred. No. 9.6e-57;
Matches 242; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 3093 caatcacgcttttagtacgctttatcaagggcgctataatcattatcgattaatggg 3152

ORIGIN	
Query Match	1.8%; Score 62.8; DB 64; Length 337;
Best Local Similarity	52.3%; Pred. No. 1 9e-07;
Matches 139; Conservative	0; Mismatches 127; Indels 0; Gaps 0;
Qy 2107	ccctctgtcagtaaaaccttctgaaagctccggagagcatcaaacagcagcattccacgcg 2166
Db 1	CCCTTCACAGCCGTAATATCTCGAGTCTCGAGGTCATCCGTCGATGGACTTCGCCACA 60
Qy 2167	ccgctcatggctcagatcatcatatttctgacatccatcccaatgccgttttgcttccggaacc 2226
Db 61	TTGACCACGCCACATCCATCATGGAGCTGCTCTCCGTTGCCGATCATCCCTTCCCAAGT 120
Qy 2227	tactctgaccattctctcgggatactttttacgaagctcgaagcagctgcgtcgag 2286
Db 121	TCCTTCGCCCCAGTCTCCCAAGTACTCTTCACCCGACAGTAGAGACACGTGAGGAAGA 180
Qy 2287	ctccggagcactcgtgtcccatcttgatgatttccatccagcagacaagaagtcca 2346
Db 181	TGTCACACGGAAGTTCGCCGCCCTCGGCTTCATCTTCATCCGATCGCGTAATTCG 240
Qy 2347	tgtcagatgcgaaagtttgagagcat 2372
Db 241	TCTCTGATCCTACCCGATTGAGGAAT 266
RESULT 12	
AW701427	
LOCUS	AW701427 506 bp mRNA EST 18-APR-2000
DEFINITION	uq86a04.y1 NCI-CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937966 5' (HUMAN); mRNA sequence.
ACCESSION	AW701427 GI:7585558
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 506)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) On Jan 6, 2000 this sequence version replaced gi:6677089. Other_ESTs: uq86a04.x1 Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
FEATURES	
source	MGI:1050386 Seq primer: -40RP from Gibco High quality sequence stop: 449. Location/Qualifiers 1..506 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:2937966" /clone_lib="NCI-CGAP_Lu33" /tissue_type="pooled lung tumors" /lab_host="DH10B (phage-resistant)" /notes="Organ: lung; Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st

Db 358 TATATRYGHW 348

Search completed: September 4, 2000, 00:24:33
Job time: 19848 sec


```
GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:48:39 ; Search time 11551.9 Seconds
        (without alignments)
        582.134 Million cell updates/sec

Title:
Perfect score: 3768
Sequence: 1 gttattaaagctgcaaatat.....aatattttacagatttaa 3768

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_bal.*
2: gb_ba2.*
3: gb_on.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pli.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pli.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_vi.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_bal.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_vil.*
82: gb_vil2.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3768	100.0	3768	33	DMU38238
2	3053	81.0	44517	42	AC018013
3	3053	81.0	273521	34	AE003515
4	853.2	22.6	57873	51	AC010021
5	714.6	19.0	92677	51	AC009382
6	524.4	13.9	57873	51	AC010021
7	462.8	12.3	92677	51	AC009382
8	166.6	4.4	4154	50	U00968
9	165	4.4	4154	5	AR059298
10	165	4.4	4154	5	122310
11	114.4	3.0	4249	9	HSU02031
12	114.4	3.0	5197	5	AR069307
13	114.4	3.0	5197	5	122319
14	112	3.0	3816	12	CGU09103
15	105.8	2.8	3115	12	RATADD1A
16	90.2	2.4	1209	12	AB017337
17	88.6	2.4	433	3	AF102873
18	88.6	2.4	4987	12	GU12330
19	82.6	2.2	1767	12	CGU22819
20	82.6	2.2	1914	12	CGU12329
21	82.6	2.2	4425	12	CGU22818
22	80.8	2.1	220895	10	HS250010
23	64	1.7	89299	32	HS17B87
24	51.6	1.4	110000	31	PFMAL4P1_2
					Continuation...
					U38238 Drosophila
					AC018013 Drosophila
					AE003515 Drosophila
					AC010021 Drosophila
					AC009382 Drosophila
					AC010021 Drosophila
					AC009382 Drosophila
					U00968 Human SREBP
					AR059298 Sequence
					122310 Sequence 37
					U02031 Human Stero
					AR069307 Sequence 53
					122319 Sequence 53
					U09103 Cricetus
					L16995 Rat adl mr
					AB017337 Mus muscu
					AF102873 Sus scrofa
					U12330 Cricetus
					U22819 Cricetus
					U12329 Cricetus
					U22818 Cricetus
					HS250010
					Z59716 Human DNA s
					AL035362 Homo sapi
					Continuation...

[illegible]

Qy	841	ccatgtacttgcctttgacgcccaocagtgtagcggcgagtcgaaccattctccaccttggtc	900
Db	841		
Db	841	CCATGTACTTCTTTGCAGCCACCGTAGCAGTCTCAACCCCATCTCCACTCTGGCTC	900
Qy	901	caccaggacaacacaggtagtgcggcgagcaagaagtgagcacactggctcgt	960
Db	901		
Db	901	CACCACGACAAGCACAGGTAGTCGGGCCAGGAAGTGCGAGTGCACACCTTGGCTCGT	960
Qy	961	cacctgcgcgtatggaagtcgaagggcaaggtacctatacaaccgggttcaaaccaagtg	1020
Db	961		
Db	961	CACCTGCCGTATGGAGTCCAGGGCAAGGTACCTATCAACCGGGTTCAACCCAAAGTGA	1020
Qy	1021	aggaagttaagcgcctcggccccacaacgcacatcgagcgcgcttatcgacctcatcaacg	1080
Db	1021		
Db	1021	AGGAAGTAAAGCGCTCGGCCACAACGCCCATCGACGGCGGTATCGCATCAACTCAACG	1080
Qy	1081	aaaagataaacgagttgaagaacctggtagtggagagcaggccaagctgaaacaagtcg	1140
Db	1081		
Db	1081	ACAAGATTACGAGTTGAAGAACTTGGTAGTGGGAGCAGGCGCAAGCTGAACAAAGTCCG	1140
Qy	1141	cagtttgcggaaatccatagacaagattcgggattctgcacgccagaatacacgatctga	1200
Db	1141		
Db	1141	CAGTTTGCGAAATCCATAGACAAGATTCCGGGATCTGCAAGCCAGAATCACGATCTGA	1200
Qy	1201	aggcagaagttgaagccctgcagagaggaagctaataaggcgcgcagcgtccaaagtgaa	1260
Db	1201		
Db	1201	AGGCAGAGTTGCAAGCGCCTGCAAGAGGAGCTAATGGCACGGCAGCGCTCCAAGGTGAAGG	1260
Qy	1261	atttaactcagctlgggcaactcggcctcgttagagcatccaagaagccgcgcgagagctgc	1320
Db	1261		
Db	1261	ATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCAAGAGCGCCGCGAGAGCTCGC	1320
Qy	1321	agaccttaccacggatgcgcggactgacgcccccacgcagcgatgaatcggatccttcgc	1380
Db	1321		
Db	1321	AGACCTTTACACCGGATGCCGACTCACGCCGCCACGACGAGTGAATCGGATCCTTCGC	1380
Qy	1381	tctgcgcattgcactcggacatctgattccgcacctaccctatggttgatccaacgcga	1440
Db	1381		
Db	1381	TCTGCCCATTGCACTCGGACATCTGFTGCCGCCATCACCTATGTTGGATGCCACGCCCA	1440
Qy	1441	gctgtagcagtggcagcagcagcaatlgaagaaccactggttgtgccagactcttatgc	1500
Db	1441		
Db	1441	GCTGTACAGTGGCAGCAGCAGCAATGAAGAACCACTGGTGGTGCCCCAGCTCTATGC	1500
Qy	1501	gcggcatggccaccactctgcctcggactctgatatgttcgcactcctggcgcg	1560
Db	1501		
Db	1501	CGGCATPGGCCACCCACTCTCGCCTCGGACTCTGATGTTTTATGTTTCGCCATCTCTGGCG	1560
Qy	1561	tcaatccctcaagaccttctccagcgcgcacctatgacagtaatgacgactctggcg	1620
Db	1561		
Db	1561	TCAATCCCTTCAAGACTTCTTCACGCGGGCCATATGACAGTAATGACCATCTTGGCG	1620
Qy	1621	acatgagcgtcaaaacacattctctcttacgacgttgaagggtgaaggttttgcgtctct	1680
Db	1621		
Db	1621	ACATGAGCGGTCAAAGACGCAATCTCTCTTACGACGTGGAAGGTGAAGGTTTCTCTCT	1680
Qy	1681	ggcagcagaagttcctggatatggctattgaacttcaactgatcttgatgttgatga	1740
Db	1681		
Db	1681	GGCAGCAGAGTTCCTGGATATGGCTATTGAACTTCCACTGTATGCTTGGATPCTTGTGTA	1740
Qy	1741	aattgcctgttacggtgatccgcagctggacgcgaacgcgacctactgcccagcaca	1800
Db	1741		
Db	1741	AATTGCTGTTTACGGTGATCCGCAGCTGGACGCCGAACAGGACGCCCTACTGCCAGACA	1800
Qy	1801	ggcagcggctgacttctatttagccaaggacgtcgtctcaugcctacgcgcggttacc	1860
Db	1801		
Db	1801	GGCAGCGGCTGACTTCTATTTTAGCCAAGGACAGTCTCTCTCAGGCTCTACGCGGTTACC	1860
Qy	1861	tcaactgtctgcatatgtttgatatgaatctaccggcgctgcgcgttgagttgaacttgc	1920
Db	1861		
Db	1861	TCAACTGCTGCATATGTTTGGATTAAAGTCTACGGCGCTCGCGTTGGAGTGTATCTTC	1920

QY	1921	agacacacgtggcagttcctctcgtttctcttccatcgccctctggtctgggtcggggtcgtg	1980
DB	1921	AGACACACGTGGCAGTTCCTTCGTTTCTTTTTCATCGCCTCTGGCTGGCTCGGCTGCTCT	1980
QY	1981	caaggcgtccgggtgggctgtttagcaaacgcgcgcagcaggaaaaaacaggcgtctggcatctg	2040
DB	1981	CAGGCGGCTCGGTTGGGCTGTTTAGCAACGCCGCCAGCAGGAACAAGCGCTTGGCATCTG	2040
QY	2041	caecgaactggccctgctcttcaaccgactgaatcaattgcaactgactcgaaatgaa	2100
DB	2041	CACGGCAACTGGCCCTGCTCTTCAACCGACTGAATCAATTGCNACTGACTGGAAATGGAA	2100
QY	2101	gccgcggtgacatgaacgcggcattatgatggcacattctcgaagcaacatggctgaagtgg	2160
DB	2101	GCCGCGGTGACATGAACGGCATATGATGSCACTATTTCGCAAGCAACATGGCTGAAGTCG	2160
QY	2161	cgcaaatctactgaacccgcgcgaacaacatctgcattcaacgtaatgacagcgtttgcgaa	2220
DB	2161	CGCAAAATCTACTGACACCGCGCGAGACCATCTGCATCCACCGTAATGACAGCGTTGCCAA	2220
QY	2221	tgaagcgcagtgcgccaaaatggttgcacagttcttccgccqatctacatgagccggg	2280
DB	2221	TGAAGCGCAGTGCCTCCCAAAATGGTTGCAACAGTCTTTCGCCCGCATACTACATGACCCGG	2280
QY	2281	ctgtcaagagtcgcggtcgcacctagggccaccgagcaaacgcgagagctacgtttgggcat	2340
DB	2281	CTGTGTCAAAGTGCGGTGCACACTAGGSCCACCAGCAAAACGCGAGGAGCTACGTTGGGCAT	2340
QY	2341	tcacagcctatggatatcgcctactcggccacgcacagctcttcacytacgatatcgagcga	2400
DB	2341	TCACAGCCCTATGGGATATCGGTACTTGGCGCCACGCACCGCTTTCACGTACGATCTGACGCACT	2400
QY	2401	ccggcgagcagatggatcttccacagctctcttaggaatccatgtgatccgctgcgccacg	2460
DB	2401	CCGCGGACGACGATGGATCTTTCACAGCTCTTAGGAATCCATGTCGATCCCGCTGCCACG	2460
QY	2461	tcattaagcaatctcgagagcatttgcgtgttaaatccattcagtgcttggttagagcgg	2520
DB	2461	TCATTAAAGCAATATCGAGAGCATTTGCTGTTTTAAATCCATTTCAGTCTCTGGTAGGAGCG	2520
QY	2521	gcccaaatcgggaaggcctgccacatctctctgcaggcgagagcggaacagttgcagc	2580
DB	2521	GCCCAAAATCGGAGGGCTTCCCCACATCTTCTGTCAGCGGAGAGCGGCAACAGTTTGCAGC	2580
QY	2581	aacagcagcacgcgcaccatctgcagcaatgttcttaagtacacgttcctcccttaagg	2640
DB	2581	AACAGCACGACGCGGCACCATTTGTCAGCAATGTTCTTAAGTACACGTCCTCTTAAAG	2640
QY	2641	acactctgggctgatgaggaatgagcggaatacaaacgtgggtggtgggcgcatgttt	2700
DB	2641	ACACTCTTGGGCTGATGAGGATGAGCGGATACAAACGCTGCTGTTGGTGGGCGCATGTTT	2700
QY	2701	tggagaccgagtacactgctctctggtagaagacacgtcggcgagcaaatgttacggca	2760
DB	2701	TGGAGACCGCATGCATCTGCTTGGTGAAGACACACGCTGGCCGACGAAATGTACGGCA	2760
QY	2761	ggatcaagcaaatgcccacgcagctgcacagtcgagcggaacacgatctctgcccaagg	2820
DB	2761	GGATCAAGCAAAATGCCCAACGCACCTGCACAGTGGCGGCAAAACCATCATCTGCCCAAG	2820
QY	2821	cgtgcattgctgtctgcgagctaaatgatcttactaaaaaaccaatggcaacgcactgg	2880
DB	2821	CGTGCATGCTGTGCTCGGAGCTAAGATGATCTTACTAAAAAACAATGGCAACGCACCTGG	2880
QY	2881	acaaaagtctcaagcaattggtaaacatcctctcgcgatgagtcgagtgctgagctccaag	2940
DB	2881	ACAAAAGTCTCAAGCAATGGTAACATCCTCTGGCATGAGTCGAGTCGAGCTGGAGCTCCAAG	2940
QY	2941	agtcttgcagtcaaccgatcacgcacqcccaagggtataaagctgccttttcagttgc	3000
DB	2941	AGTCTTGCATGCTCACCGGATCACCGACGCCCAAGGGTATAAAGCTGCTTTTTCAGTTGC	3000
QY	3001	ttacctgcgattgctgctcqaacctaagdaactgctctctgttggaaacttgaaacacatgaata	3060

```
|||||
Db 3001 TTACCTGGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACCTGGAACACATGAATA 3060
Qy 3061 tggagacgagtgcttctaccagtgccagtggaagtgcgagaagttccagaccgatt 3120
Db 3061 TGGAGGACGATGGCTTCTACCAAGTGCAGGTGAAGTGCAGAAAGTTCAGACCGATT 3120
Qy 3121 tgaactcgttgcgaacatgtgagataataccgaacgcccatacgcgcatatttgt 3180
Db 3121 TGAACCTGTTGCGCAACATGTGGAGAAATATACCGAACGCCAATCGCGCATATTTGT 3180
Qy 3181 acgaggcagttgtcgctgatggctgagcctcaccggtgtccaaacgaacagctcttg 3240
Db 3181 ACGAGGAGTTGTGCGCTGATGGCTGAGCTCACCGCTCACCGTGTCCAAACGCAACAGCTCTGG 3240
Qy 3241 acaggagctcgcgatacgcgaacgcccactcgtccatctcttcgcccagcagaagatcgc 3300
Db 3241 ACAGGAGTCTGCGATACGCAAGCCGACCTCGTCCATCTTCTGCGGAGCAAGATCGGC 3300
Qy 3301 ggcagcaaacctcgtggtgagagcggaacggggttcggccatgtacgtggtcgtgca 3360
Db 3301 GGCAGCAAACTCTGTCGGCGGAGAGCGGGAACGGGCTTCGGCCATGTACGTGGCTGCA 3360
Qy 3361 agtatcccgctgcgtcgtcagctcccccgggtggaacggtgctggtgcatgttagcggg 3420
Db 3361 AGTATCTCCGCGCTGCTGCTCAGCTCCCGGGTGAACGTGCTGGCATGTTAGCCGAGG 3420
Qy 3421 cggccaaagacctgagaagtggtggtgagcgaagcgaagcgaagcgaagcgaagcgtga 3480
Db 3421 CGGCCAAGACCTTGGAAGAGTGGCGGACCAAGCGAAAGCTCAAGGAGTGTACAGCTGA 3480
Qy 3481 tgaagtcgctgggcaacgagcgtggtgagcgtggaagcgttaggatagtagtgaaagacata 3540
Db 3481 TGAAGTCGCTGGGCAACGGCATTGGCAGCGTGAAGGCTTAGGATAGTAGTGAAGTACATA 3540
Qy 3541 ataagtggaacgagcgtggtggtgatttcagcaaatgaatacccggttgcattcaaaa 3600
Db 3541 ATAAGTGGCAGCAACGCTGGTGTGATTTTCAGCAAAATGAATACCCGTTGCTATTCAAAA 3600
Qy 3601 gaattacaatgctaggtcttataattacgctatctctctgttttccagcccggtta 3660
Db 3601 GAATTAACAATGCTAGTGTCTTTATAATTACGCTATTTCCTGTGTTTTCCAGCCCGGTTA 3660
Qy 3661 tgettagattgaattttaaattatttaattgacattttattttattattattacc 3720
Db 3661 TGCTTAGATTGTAATTTTAAATTTTAAATATGATGACATTTTATTGTTATTATTACC 3720
Qy 3721 gtacttgttaaacgtatttataacaataaatttttaacagatttaa 3768
Db 3721 GTACTTGTAAACGTATTATTATAACAATAAATATTTTAAACAGATTAAA 3768

RESULT 2
AC018013
LOCUS AC018013 44517 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC018013
VERSION AC018013.1 GI:6553177
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 44517)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213012 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
```

```
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source              Location/Qualifiers
    1..44517
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
BASE COUNT 12813 a 9534 c 9565 g 12605 t
ORIGIN

Query Match      81.0%; Score 3053; DB 42; Length 44517;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 3594; Conservative 0; Mismatches 5; Indels 488; Gaps 5;

Qy 170 agatttcgaatgagacagacactgatgaacttaataagacgctccgctggacagtcacat 229
Db 18925 AGCATTCGCAATGACACGACACTGATGAACCTTAATAGACGCTCCGCTGGACGAGTCCAT 18984
Qy 230 ggaattgttcaaaagcgagagatgtcttcgaacgcttcgacgacgacgacgacgacat 289
Db 18985 GGATTTGTTCAAGCGGAGGATGTCTTCGAACCGTTTCGACGCCGACCTGCACTCGGACAT 19044
Qy 290 gctggacatcatctcaacgacatggacctggcgccgacgacgacatgtacaacatgctgct 349
Db 19045 GCTGGACATCATCTCAACGACATGGACCTGGCGCCGACGACGATGTACAACATGCTGCT 19104
Qy 350 ggaagagcctcgaacgacataccagcagacgacgacgacgacgacgacgacgacgacgac 409
Db 19105 GGACGAGCCTCGAAGCGATACCCAGCAGACGACGCTCGTGGATCAGACGCCGCAATCCGT 19164
Qy 410 cgagcaacagccacagcgtgaaagcgagcactcttcgacgacgacgacgacgacgacgac 469
Db 19165 CGAGCAACAGCCGACGCTGAAGACCGAGCAGCTCTTCGCCAGTGCACATCAAGGAGGAAT 19224
Qy 470 gcatcagcagcaacacagtcgcgcttctcgtctatacaaacagatccctcatagccac 529
Db 19225 GCATCAGCAGCAACACAGTCGCGCTTCTCGTCTACAAACACAGATCCCTCTCATAGCCAC 19284
Qy 530 aagctacaattgtccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 589
Db 19285 AAGCTACAATTGTCCCCAGCAACAGCCGACGCGGCGCTTTTGAAGGCCGCCCAACCAACAGC 19344
Qy 590 caccatacatcaatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 649
Db 19345 CACCATACATCACATGACGCCGCCAGCGGATGCGCGCGAAGACAGCGCGGTGTATCCCCATC 19404
Qy 650 tctgggagtagtcttctaccagtcacgtctccccgcccacgctcgcgctggagtgctgc 709
Db 19405 TCTGGCAGTAGCTTTTGTCTACCACTCCATGTCCCCGCCACGTCGCGCGGTGGAGTCTGC 19464
Qy 710 gaaccagaatgcaatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 769
Db 19465 GAACCAAAATGTCAATGTCAATGACGCCCGTGTGTGCAACTCTCTGCTCCCGCTTCTGCTCC 19524
Qy 770 ttgcccacagcagtcgtagtcgcaaaccttcattacgtacaactctaagccggaatgac 829
Db 19525 TTTGCCCCAGCAGTCGTATCGCAACCCCTTCATTACGTACAACCTTAAGCGCGGAATGAC 19584
Qy 830 ttcogataagagcagtgactgtctcttgccgcccacgagtgagcagtcaccccccatctcc 889
Db 19585 TTCCGATGAAGCCATGTACTTCTCTTCGACGCCACCGGTAGCCAGTCCCAACCCCATCTCC 19644
Qy 890 acctgtgctccaccacgacagcacagtagtcgggcccagcaagtgcgagtgggcacc 949
Db 19645 ACCTGTGGCTCCACCAACCGACAGCAGGATGTGCGGCGCAGCAAGGTGGAGTGGGACCC 19704
Qy 950 actggtccctcactcgcgctatggaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaag 1009
Db 19705 ACTGGCTCCGCTACCTGCCGCTATGGAAGTCCAGGGCAAGGTACTTATCAACCGGGTTCA 19764
Qy 1010 acccaagtgtaaggaagtaagcgtcgtcgcccaacacgcccacatcgagcggcgtatcgac 1069
|||||
```


QY 2987 ----- 2986
Db 21985 ACTCTTCCCTGGAGATGACACGAAGTGCGATTGTAGTCCCATTTGTCTGTGCCAATTGCT 22044
QY 2987 ----- 2986
Db 22045 TCAGGCAATACAACTTCAGTGCAAGCTCTGTGCATTTTCATACACAAGCGCGAATGTCTCT 22104
QY 2987 -----gcttttcc 2994
Db 22105 GGCAAAGTGCACCAAGTTTAAAGCGCCACTAATCATATTATTTCTTATCGACGCTTTTCC 22164
QY 2995 agttgcttaactgcgattggtgctgcgaactaggactgctctgtggaactggaacaca 3054
Db 22165 AGTTGCTTACCTGGATTGGCTGCTCGAACTAGGACTGCTCTGTGGAACTGGACACA 22224
QY 3055 tgaatagggagcagatggtgttcttaccgaagtgcgaagtgtaagtgctcgagaagtccaga 3114
Db 22225 TGAATATGGAGGACGATGGCTTCTACCAAGTGCAGGTGAAGTCTCGAGAAGTTCGAGA 22284
QY 3115 ccgatttgaaactgttgcgcaacattgtggaatatacc----- 3154
Db 22285 CCGATTTGAACCTGTTGCGCAACATTTGTGGAGAATATACCGGTATATGCAAAATCCTTACC 22344
QY 3155 -----gaacgcc 3161
Db 22345 CTTCCAATATGTCTGTTCCGTAACCTCAGCTTGTGTAAACTTTTGCCATTACAGACGCC 22404
QY 3162 caatcgcgcatatattgttacgaggcagttgttcgctgtgtagctgagcctcacgctgt 3221
Db 22405 CAATCGCGCATATATTGTACGAGGAGAGTTTGTGCGCTGATGCTGGAGCCTCACGCTGT 22464
QY 3222 ccaacgcaacagctcttggacagagctctcgatcacgcaaccccaactgcctcatcttc 3281
Db 22465 CCAACGCAACAGCTCTTGGACAGAGTCTCGATCATCGCAACGCCCACTCGTCCATCTTC 22524
QY 3282 tgcggcagcaagatgcggcgagcagaaacttcgtgtggcgagagcggaacgggcttcg 3341
Db 22525 TCGCGCAGCAAGGATCGCGGACGAGCAACTTCTGTGGCGGAGAGCGGAAACGGGCTTCG 22584
QY 3342 gccatgtacgtgacctgaagtattctccgctgcgctgctgaagctcccggtgaacgt 3401
Db 22585 GCCATGTACGTGGCCCTGCAAGTATCTCCCGCTTCGCGTCTCAGCTCCCGGTGAACGT 22644
QY 3402 gctggcatgttagccgagcgcccaagaccctggagaagtggtggcgacaaagcgaagctc 3461
Db 22645 GCTGGCATGTGTAGCCGAGGCGGCCAAGACCCTGGAGNAGTGGCGNCAAGCCAAAGCTC 22704
QY 3462 aaggagtgtaccagctgtatgaagtcgctggcgcaacggcattggcgagcgtgaaggcttag 3521
Db 22705 AAGGAGTGTCTACCAGCTGTATGAAGTCGCTGGCAACGGCATTTGGCAGCGTGAAGGCTTAG 22764
QY 3522 gatagtagtaagtacataataaagtgagcagcaagcgtggtgtgtagattttcaagaaatgaat 3581
Db 22765 GATAGTAGTGAAGTACATAATAAGTGGCAGCAACGGTGTGGATTTCAGCAAAATGAAT 22824
QY 3582 acccgttgtctatcctaaagaatacaaaatgcctaggtctctttataattacgctattcctc 3641
Db 22825 ACCGTTTGTATTCAAAAGAAATACAAATGCCTAGTCTCTTTATAATTAGCTATTCCTC 22884
QY 3642 tgtttccacgcccgttatgcttagattgtaatttttaaaattatttaattatgacattt 3701
Db 22885 TGTTTTCCACGCCCGTTATGCTTAGATTGTAATTTTAAATTTATTAAATATGACATTT 22944
QY 3702 tatttgtttatttaccgtactgtttaacgtattttataacaaataaataattttaacag 3761
Db 22945 TATTGTTTATTATTACCGTACTTGTAAACGCTATTATTATAACAATAAATATTTTTAACAG 23004
QY 3762 atttaaa 3768
Db 23005 ATTACA 23011

RESULT 3
AE003515/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
FEATURES
source
mrna
gene

AE003515 273521 bp DNA INV 22-MAR-2000
Drosophila melanogaster genomic scaffold 14200001386050 section 2
of 54, complete sequence.
AE003515 AE002602
AE003515.1 GI:7293706
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 273521)
Adams,M.D., Celiniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amantides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Basu,A.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cavley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrier,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Idegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kernison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacle,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J.D., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirska,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 273521)
2 273521
Adams,M.D., Celiniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
Location/Qualifiers
1. .273521
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
Join(<2612..>22692,22755..>22918,23186..>23319,23379..>23449,
23505..>23597)
/gene="CG14184"
/product="CT33793"
/db_xref="FLYBASE:FBgn0014184"
/db_xref="FLYBASE:FBgn0036932"
/evidence="not_experimental
<22612..>23597

QY	350	ggacgagcctcgaaacacataccacagacagcagctcggtgagatcagcagcgcgaatccggt	409		
Db	256210	GGACGAGCCTCGAACCATACCCAGACAGCGAGTCGGTGGATCAGACGCCGAATCCGT	256151		
QY	410	cgagcaacgcccgcagctgaaagcagagcaactcttcgccagtgacacatacaagaggaact	469		
Db	256150	CGAGCAACAGCCGACGCTGAAAGCGAGCACTCTTCGCCAGTGCATCAAGAGGAACCT	256091		
QY	470	gcatacagcagaacaacagtcgccgtctctgctctcaaacacagatccccctcatagcaac	529		
Db	256090	GCATAGCAGCAACAACAGTCGCCGTCTCTGCTTACAACACAGATCCCCCTCATAGCCAC	256031		
QY	530	aaqctacaattgtcccccagcaacgagcgcggtctttgaaaggcgcgcacaaacacagc	589		
Db	256030	AAGTACAATTGTCCCGACCAACAGCGGAGGGCCCTTTTGAAGGCCGCCCAACCAACAGC	255971		
QY	590	caaccaacatcacatggaagcccccagcggatgccgcggaacacagcgcggtgtatcccccatc	649		
Db	255970	CACCATACATCACATGGAGCCGCCAGCGGATGCCGCCGAACACAGCGCGTGTATCCCCCATC	255911		
QY	650	ctctggcagtagcttctgtctacagtcgatccatgtcccccacagtcgcggtggagtctgc	709		
Db	255910	TCGCGCAGTAGCTTTGTCTTACCAGTCCATGTCCCGCCACAGTCGCCGCTGGAGTCTGC	255851		
QY	710	gaaccagaatgtcaatgtcatgcagcccgttgtgtgaactcctgtcccgcctctctgtccc	769		
Db	255850	GAACAGAATGTCAATGTATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCCTGCTCC	255791		
QY	770	tttgccccagcgtcgatctccgcaaccccttcattagctacaaactctaaagccgggaatgac	829		
Db	255790	TTTGGCCCCAGCAGTCGTATCCGCAACCCCTTCAATTAGTACAACCTTAAGGCCGGAATGAC	255731		
QY	830	ttccggtgaagccaatgacttctgttcagcccacggttagcagtcocaaaccccatctccc	889		
Db	255730	TTCCGATGAAGCCATGACTTCTCTTGACAGCCCGTAGCCAGTCCAAACCCCATCTCC	255671		
QY	890	acctgtggctccaccaccgaagcacaggtagtcgggcagcaaggtgcagtggcac	949		
Db	255670	ACCTGTGGCTCCACCACCGACAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACC	255611		
QY	950	actggctccgtccactcgcgctatggaagtccagggaaggttacctatacaacccgggttca	1009		
Db	255610	ACTGGCTCCGCTCACTCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTCA	255551		
QY	1010	accgaagtgaaaggaagtaagcgtcgcaccaacgcacatcgagcgcgtatcgac	1069		
Db	255550	ACCGAGGTGAAGGAAGTAAGCGCTCGGCCACACGCCATCGAGCGGCTATCGCAC	255491		
QY	1070	ctcaatcaacgacaagattaacagagttgaagaacttggtagtgaggagcagggccaagct	1129		
Db	255490	CTCAATCAACGACAAGATTAAACGAGTTGAAGAACTTGGTAGTGGGAGACAGGCCAAGCT	255431		
QY	1130	gaacaagtcgcagtgctcgggaaatccatagacaagattcgggatctgcgaacgccagaa	1189		
Db	255430	GAACAAGTCCGCAAGTGTGCGGAAATCCATAGACAAGATTTCGGGATCTGCAACGCCAGAA	255371		
QY	1190	tcacgatctgaagcgagagttgcagcgcctgcagagggagcttaatggcagcgagcgtcc	1249		
Db	255370	TCACGATCTGAAGGAGAGATTGACGCGCTTCGACAGGGAGCTAATGGCACGCGCGGCTC	255311		
QY	1250	caaggtgaaggatttacttcagctgggcactcgccctggttagagcatccaaagaagcgcgc	1309		
Db	255310	CAAGGTGAAGGATTTACTTCAGCTGGCACCTCGGCCCTGTAGAGCATCCAAGAAGCGCCG	255251		
QY	1310	cgagagctcgacagaccccttacacagatgcccggagctgaagcccccagcagcgatgaatc	1369		
Db	255250	CGAGAGCTCGCAGACCTTTACCAACGGATGCCGACTGAGCGCCGCCACGACGGATGAATC	255191		
QY	1370	gaatccttcgctctcccccacatcgacatctcgttgcgcgcacatccctctatgttg	1429		
Db	255190	GGATCCTTCGCTCTCGCCCATGCATCGGACATCTCTGTCGCCCCATCACCTTATGTTGG	255131		
QY	1430	atccaccgcagctgtagcagtggcagcgagcagcaaatgaagaaccactggtgtgcc	1489		
Db	255130	ATCCACCCGCCAGCTGTAGCAGTGGCAGCAGCAATGAAGAACCACTGTGTGTGCC	255071		
QY	1490	cgactctatgcggcgcatggccaccacactctcgctcggaactctgactgttatgttcgc	1549		
Db	255070	CAGCTCTATGCGCGGATGCCACCACCTCTGCCCTCGGACTCTGCATGTTTATGTTCGC	255011		
QY	1550	catcctggccgtcaatccccctcaagaccttctccagcgcgcgcactatgacagtaatga	1609		
Db	255010	CATCTTGGCCGTCAATCCCTTCAAGACCTTTCTCCAGCCGGCCACTATGACAGTAATGA	254951		
QY	1610	cgatcttggcgacatgagcgggtcaagagcgcattctctcttacgacgtgga	1660		
Db	254950	CGATCTTGGGACATGAGCGGTCAAGAGCGCATCTCTCTTACGAGTGGGAAGGTAGCTA	254891		
QY	1661	-----	1669	-----	aggtgaag
Db	254890	GCCTAGCATCCAGATGTTGTTAAATATTGCTAATGGGTTTCTCTTGATTTAGGTGAAG	254831		
QY	1670	tittgctgtctgcagcagagttccctggatatggctatigaaacttcacactgatcttgg	1729		
Db	254830	TTTTGCTGTCTGCGACAGAGTTCCTGGATATGGCTATTGAACTTCACACTGATGCTGG	254771		
QY	1730	atgcttgggtgaattgtctgttacggtgatccgcgagctggagcgcgcgaacgagcgccta	1789		
Db	254770	ATGCTTGGTGAATTTGCTGTTTACGGTGTATCGCAGCTGGAGCGCAACAGGACGCCTA	254711		
QY	1790	ctgcacagcaagcagcgggctgaactctctatttagccaaggaagcagctcgtctcagcccta	1849		
Db	254710	CTGCCAGCACAGCAGCGGGCTGACTTCTATTTTACCAAGGACAGCTCTCTCAGGCCCTA	254651		
QY	1850	cgccggttacctcaactgtctcatagttttgattaaactctaccggcgtcgcgcttga	1909		
Db	254650	CGCCGTTTACCTCAACTGCTGCATATGTTTGGATTAAGTCTACCGCGCTCGCGCTTGA	254591		
QY	1910	gtgttacttgcagaccacgtggcagttccttcttcttcttccatcgccctctggctggg	1969		
Db	254590	GTGTTACTTGCAGACACCGTGGCAGTTCCTTCGTTTCTTTTCCATCGCTCTGTGGTGG	254531		
QY	1970	tcgggtgctgtcaagcgggtccggtaggctgtttaagcaacgcgcgcagcaggaacagcc	2029		
Db	254530	TCGGGTGCTGTCAAGCGGTCCGGTGGGTGTTTACCAACGGCGCGCAGGAAACAGGC	254471		
QY	2030	gctggcatctgaacgcgaactggcctgctcttcaacccagctgaatcaattgcaactgac	2089		
Db	254470	GCTGGCATCTGCACGGCAACTGGCCCTGCTCTTCAACCCGACTGAATCAATTGCAACTGAC	254411		
QY	2090	tggaaatggaagccgcggtgacatgaacggcattatgtatggcactattcgcaagcaacat	2149		
Db	254410	TGGAATGGAAGCGCGGTGACATGAACGGCATTTATGATGGCACTATTGCAAGCAACAT	254351		
QY	2150	ggctgaagtggcgacaaatctactgaacgcgcgcgagacacatctgcacatgaatgac	2209		
Db	254350	GGCTGAAGTGGCGCAATCTACTGACACCGCGGAGACCATCTGCATCCACGTAATGAC	254291		
QY	2210	agcgttgcgaatgaagcgcagtgcccccataatggttgcaacagttcttcgcccgatacta	2269		
Db	254290	AGGTTGCCAATGAAGCGCAGTGCCTCCCAAAATGGTTGCAACAGCTCTTCGCCCGATACTA	254231		
QY	2270	catgagccgggctcgtcaagagtgcggtcgcaatagggccacccagcaaacgcgagagct	2329		
Db	254230	CATGAGCCGGGTCTGTCGAAGAGTGGCGTCTCTACTAGGGCCACCGAGCAACGAGAGCT	254171		
QY	2330	aggttgggattcaacagctatgatatcgctactgcgcagcagcagcttctcagctacga	2389		
Db	254170	AGTTGGGATTTACAGACCTATGGATATCGTACTCGGCCAGCAGCTCTTTCACGTACGA	254111		
QY	2390	tctgagcactccgcgcagcagatggattcttcacagctcttaggaatccatgtatcc	2449		
Db	254110	TCTGAGCGACTCCGGGAGCAGGATGGATTCTTCACACGCTCTTAGGAATCCATGTATCC	254051		
QY	2450	cgctgccacgtcatta-----	2466		

Db 254050 CGCTGCCACGTCATTAAGTTAAGTTGGACTGAATTTCTAGTGTATCGCACAAATTCAA 253991
Qy 2467 -----agcaataatcgagagcatttgcgtgtttaaataccat 2500
Db 253990 AACTAATTGTTCTTTTCCAAATAATCAGCAATATCGAGACATTTGCTGTTTAAAGCCAT 253931
Qy 2501 tcagtgtcttggttaggagcgggcccacaaatcgggagggcctgcccacatacttctgtcagcgg 2560
Db 253930 TCAGTGTCTGGTAGGAGCGGGCCACAAATCGGAGGCCCTGCCACATCTTCTGTCAAGCG 253871
Qy 2561 agaggcgaaacagttgcagcaacacagcagcacagcggcacacattgtcaacaatttcttaa 2620
Db 253870 AGAGCGGAACAGTTGTCAGCAACAGCAGACAGCGGCACCATTTGTGAGCAATGTTCTTAA 253811
Qy 2621 gtcacgctccctcttaaggacactcttgggctgagtgcagtgagcgagacatacaacgt 2680
Db 253810 GTACAGGTCCTTCTTAAAGACACTCTCTGGGCTGATGAGGATGAGCGGATACAAACGT 253751
Qy 2681 gdtgtggtggccgagtttttgagaccgcagtgcaactggctcccttggtgaaagacacgct 2740
Db 253750 GGTGTGTTGGCCGATGTTTGGAGACCGCAGTGCACCTGGCTTGGTGAAGACACGCT 253691
Qy 2741 gcccagcaaatgtcagcgagagatacaagcaaatgcccacgcagctgcacagtgcgga 2800
Db 253690 GGCCGAGCAATTGTACGGCAGGATCAAGCAAAATGCCACGCCAGCTGCAACAGTGGCGCA 253631
Qy 2801 aaacatcatctgccaaagcgctgcatactgtgcagctgcagctaaagataatctactaaa 2860
Db 253630 AAACGATCATCTGCCCAAGGCGCTGCATCTGTCTGCCAGCTAAGATGATCTTACTAAA 253571
Qy 2861 ----- 2860
Db 253570 GTGAGTAAACTAGAACAACTAATGAAAGAACAAACAACTAATCATCGCACACTTTTGCAG 253511
Qy 2861 aaacaatggcaacgcactggacaaagtctcaagcaatggttaaacatcctctgogatga 2920
Db 253510 AAACAATGGCAACGCCACTGGACAAAAGTCTCAAGCAATTTGGTAAACATCCTCTGCCATGA 253451
Qy 2921 tgcagagtggagctccaaagtgcttaactatcactgcagcgatcacccagccaagggat 2980
Db 253450 GTCGAGTGTGGAGCTCCAAGAGTGCTTGACTGTCAACCGGATCACCGCAGCCAAAGGGTAT 253391
Qy 2981 aaagct----- 2986
Db 253390 AAAGCTGGTAAGAAATGTGTGCACCCCGAACAACTAGTTTTTGGCGACTGAGTCCCACCAG 253331
Qy 2987 ----- 2986
Db 253330 ACTCTTCCCTCTGGAGATGACACGAAGGTGCATTGTAGTCGCAATTGTCTGTCCCAATTGCT 253271
Qy 2987 ----- 2986
Db 253270 TCAGCAATACAACTTCAGTGCAGCTGTGTGCATTTACATACAAACGCCGAATGTCCT 253211
Qy 2987 -----gcttttcc 2994
Db 253210 GCCAAAGTCGCCAAGTTTAAAGCGGCACATAATCATTTATTTCTATGCGACTTTTCC 253151
Qy 2995 agttgcttacctgcgattggtctgcgaacttaggactgctctgtgggaactggaaacaca 3054
Db 253150 AGTGTCTTACCTGCGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACCTGGAACACA 253091
Qy 3055 tgaatatggagagacatggctctaccagtgccagggtgaaagtgctcagaaagttccaga 3114
Db 253090 TGAATATGGAGACGATGGCTTCTACCAAGTCCCAAGTCCAGGTGAAGTGTCTCGAAGATTTCCAGA 253031
Qy 3115 ccgatttgaactcgttcgcgaacatttggagaatatcc----- 3154
Db 253030 CCGATTTGAACTCGTTGGCCACACATTTGGGAGAAATATACCGGTATATGCAAAATCCTTTACC 252971
Qy 3155 -----gaacgcc 3161
Db 252970 CTTTCCAATATGTCGTTCCGTAACTGAGCTTGTGTAAACTTTTTCGCATTTACAGAACGCC 252911

Qy 3162 caatcgcgcataatatgttacgagcgagtttgcgctcgtgatggctggagcctcacctgt 3221
Db 252910 CAATCGGCATATATTGTACGAGGCAGTTTTCGCCCTGATGGCTGGAGCCTCACCGTGT 252851
Qy 3222 ccaacgcaacagctcttgacagagatctcgcgatcacgcaacgcccaactcgtccatcttc 3281
Db 252850 CCAACGCAACAGCTCTTGGACAGGAGTCTCGGATCACGCAACGCCCACTCGTCCATCTTC 252791
Qy 3282 tgcggcagcaagatcggcgcgagcagaacttcgtggcgaggaagcgaggaacggggcttcg 3341
Db 252790 TGGCGGACGAAGATCGGGCGACAGCAACTTCGTGGCGCGAGAGCGGAACGGGCTTCG 252731
Qy 3342 gccatgaogtgcctgcgaagtatctcccctcgtcgtcagctcccgggtgaaacgt 3401
Db 252730 GCCATGTACGTGGCCTGCAAGTATCTCCGCCCTGCGGTGCTCACTCCCCGGGTGAACGT 252671
Qy 3402 gctggcatgttagccgagcgcccaagacctggagaagtgggcgagcaagcaaaagctc 3461
Db 252670 GCTGGCATGTTAGCGAGCGCGCCAAGACCTTGGAGAAGCTGGGCGACAAGCAAGCTC 252611
Qy 3462 aaggagtgctacagctgatgaaagtcgctgggcaacggcaatggcagcgtgaaaggcttag 3521
Db 252610 AAGGAGTGTCTACGAGCTGATGAAGTCCGTGGCAACGGCATTTGGCAGCGCTCAAGGCTTAG 252551
Qy 3522 gatagtagtaaatcatataaagtgcgcacgaacgtggtgtgtaatttcacaaaatgaat 3581
Db 252550 GATAGTAGTAAAGTACATAATAAGTGGCAGCAACGCTGGTGTGGATTTTCAGCAAAATGAAT 252491
Qy 3582 accggttgcattcctaaagaattacaatgcctagctcttataataacgctatttcctc 3641
Db 252490 ACCGGTTTCTATTCAAAAGAAATTACAAATGCCTAGGCTTTTAAATTACGCTATTTCCTC 252431
Qy 3642 tgtttccacgcgcggttatgcttagattgtaattttataaattatttaatatggacattt 3701
Db 252430 TGTTTTCCACGCCCGGTATTGCTTAGATTGTAATTTTAAAAATTATTTAATATGGACATTT 252371
Qy 3702 tattgtttattattaccgtacttgttaacgtatttataacaataattatttaacag 3761
Db 252370 TATTTGTTTATTATTATACCGTACTTGTAAACGTAATTTTATAACAATAAATATTTTTAACAG 252311
Qy 3762 atttaaa 3768
Db 252310 ATTACA 252304

RESULT 4
AC010021/c
LOCUS
DEFINITION
AC010021 57873 bp DNA HTG 18-FEB-2000
Drosophila melanogaster clone RPC198-11A8, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.
AC010021
AC010021.3 GI:6996706
HTG: HTGS, PHASE1.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 57873)
Muzny, D. M., Adams C., Bailey, M., Barberia, J., Blankenburg, K.,
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K. J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J. H., Gorrell, L. L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,
Hollway, C., Hosak, H., Jackson, L. E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, P., Li, Z.,
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R. J., Lu, J.,
Lucier, R., Martin, R., Martinez, C., McLeod, M. P., Mei, G., Morgan, M.,
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L. L.,

Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Suggang, R.,
 Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbah, M.,
 Watlington, S., Weinstock, G., Weinstock, I. R., Williamson, A.,
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D., and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 57873)
 Worley, K.C.
 Direct Submission
 Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 18, 2000 this sequence version replaced gi:5881473.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

 Center project name: DRBU
 Center clone name: RPC198-11A8
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy: 99% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 25215 bases at least Q40
 Consensus quality: 39511 bases at least Q30
 Consensus quality: 46436 bases at least Q20
 Estimated insert size: 45078; sum-of-contigs estimation
 Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 46 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 831: contig of 831 bp in length
 * 851: gap of unknown length
 * 852 1687: contig of 836 bp in length
 * 1688 1707: gap of unknown length
 * 1708 2546: contig of 839 bp in length
 * 2547 2566: gap of unknown length
 * 2567 3384: contig of 818 bp in length
 * 3385 3404: gap of unknown length
 * 3405 4818: contig of 1414 bp in length
 * 4819 4838: gap of unknown length
 * 4839 5673: contig of 835 bp in length
 * 5674 5693: gap of unknown length
 * 5694 6598: contig of 905 bp in length
 * 6599 6618: gap of unknown length
 * 6619 7540: contig of 922 bp in length
 * 7541 7560: gap of unknown length
 * 7561 8940: contig of 1380 bp in length
 * 8941 8960: gap of unknown length
 * 8961 10288: contig of 1288 bp in length
 * 10289 10268: gap of unknown length
 * 10269 11351: contig of 1083 bp in length
 * 11352 11371: gap of unknown length
 * 11372 12481: contig of 1110 bp in length
 * 12482 12501: gap of unknown length
 * 12502 14007: contig of 1506 bp in length
 * 14008 14027: gap of unknown length
 * 14028 14865: contig of 838 bp in length
 * 14866 14886: gap of unknown length
 * 14886 16199: contig of 1314 bp in length
 * 16200 16219: gap of unknown length
 * 16220 17277: contig of 1058 bp in length
 * 17278 17297: gap of unknown length

17298 * 18246: contig of 949 bp in length
 18247 * 18266: gap of unknown length
 18267 * 19687: contig of 1421 bp in length
 19688 * 19707: gap of unknown length
 19708 * 21218: contig of 1511 bp in length
 21219 * 21238: gap of unknown length
 21239 * 22141: contig of 903 bp in length
 22142 * 22161: gap of unknown length
 22162 * 23043: contig of 882 bp in length
 23044 * 23083: gap of unknown length
 23084 * 23479: contig of 415 bp in length
 23480 * 23498: gap of unknown length
 23499 * 25055: contig of 1557 bp in length
 25060 * 25075: gap of unknown length
 25076 * 26050: contig of 975 bp in length
 26051 * 26070: gap of unknown length
 26071 * 27618: contig of 1548 bp in length
 27619 * 27638: gap of unknown length
 27639 * 28649: contig of 1011 bp in length
 28650 * 28669: gap of unknown length
 28670 * 29958: contig of 1289 bp in length
 29959 * 29978: gap of unknown length
 29979 * 30893: contig of 915 bp in length
 30894 * 30913: gap of unknown length
 30914 * 32447: contig of 1534 bp in length
 32448 * 32467: gap of unknown length
 32468 * 33518: contig of 1051 bp in length
 33519 * 33538: gap of unknown length
 33539 * 34583: contig of 1045 bp in length
 34584 * 34603: gap of unknown length
 34604 * 35754: contig of 1151 bp in length
 35755 * 35774: gap of unknown length
 35776 * 37161: contig of 1387 bp in length
 37162 * 37181: gap of unknown length
 37182 * 38633: contig of 1452 bp in length
 38634 * 38653: gap of unknown length
 38654 * 39940: contig of 1287 bp in length
 39941 * 39960: gap of unknown length
 39961 * 41986: contig of 2026 bp in length
 41987 * 42006: gap of unknown length
 42007 * 43675: contig of 1669 bp in length
 43676 * 43695: gap of unknown length
 43696 * 44832: contig of 1137 bp in length
 44833 * 44852: gap of unknown length
 44853 * 46381: contig of 1529 bp in length
 46382 * 46401: gap of unknown length
 46402 * 47470: contig of 1069 bp in length
 47471 * 47490: gap of unknown length
 47491 * 49244: contig of 1754 bp in length
 49245 * 49264: gap of unknown length
 49265 * 50791: contig of 1527 bp in length
 50792 * 50811: gap of unknown length
 50812 * 52579: contig of 1768 bp in length
 52580 * 52599: gap of unknown length
 52600 * 53959: contig of 1360 bp in length
 53960 * 53979: gap of unknown length
 53980 * 56252: contig of 2273 bp in length
 56253 * 56272: gap of unknown length
 56273 * 57873: contig of 1601 bp in length.
 ----- Location/Qualifiers
 1. .57873
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RPC198-11A8"
 BASE COUNT 15592 a 12473 c 12485 g 16366 t 957 others
 ORIGIN
 Query Match 22.6%; Score 853.2; DB 51; Length 57873;
 Best Local Similarity 89.8%; Pred. No. 1.1e-192;
 Matches 965; Conservative 0; Mismatches 39; Indels 71; Gaps 3;
 QY 1858 acctcaactgtctgcataatgttggattaaagtaccgcgtgcgttgagtggtact 1917

```
|||||
Db 47494 ACCTNNNNNNNNNNNNNNNNNNNGAATTAAAGTCTACCGCGCTCGCGCTTGGAGTGTACT 47435
Qy 1918 tqcacacacgtggcagtccttcgtttcttcttccatcgctctggctgggtgc 1977
|||||
Db 47434 TGCAGACCAGTGCAGTGTTCCTTCGTTTCTTTCATCGCCTCCCTGGCTGGGTGGTGC 47375
Qy 1978 tgcacgcggtccggtggctgttagcaacgcccgcagcaggaaacaggcgctggcat 2037
|||||
Db 47374 TGTACGGCGGTCCGGTNGCTGTTTACGAACGCCAGCAGGAACAGAGCGCTGGCAT 47315
Qy 2038 ctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgactggaatg 2097
|||||
Db 47314 CTGCACGGCACTGGCCCTGCTCTTCAACCGCACTGAATCAATTGCAACTGACTGGAAATG 47255
Qy 2098 aaagccgcggtgacatgaacgcgcatatgatgagcactattccaaagcaacatggctgaag 2157
|||||
Db 47254 GAAGCCGGGTGACATGAACGCATTTATGATGGCACTATTCCGAAGCAACATGGCTGAAG 47195
Qy 2158 tggcgcaaatctactgacacccgcgcgagaccatctgcatccacgtaaatgacagcgttgc 2217
|||||
Db 47194 TGGCCCACAATCTACTGACACCGCGGAGACCATCTGCATCCACGTAAATGACAGCGTTCC 47135
Qy 2218 gaatgaagcgagtgcccccaaatggttgcaacagttcttcgccgatactacatgagcc 2277
|||||
Db 47134 GAATGAAGCGCAGTCCCCCAAAATGTTGCAACAGCTCTTCGCCCGATACTACATGAGCC 47075
Qy 2278 gggctcgtcaagagtgcgctgcactaaggccacgcagcaaacgagcaggtacgttggg 2337
Db 47074 GGGCTCGTCAAGAGTGCGGTCTCATAGGGCCACCGCAACGAGGAGTACGTTGGG 47015
Qy 2338 cattcacagcctatgatgctactgcgcacgcacgcttcttcgctacgactgagcg 2397
Db 47014 CATTCACAGCCTATGATATCGCTACTGCGCCACGCACGCTCTTCAGTACGATCTGAGCG 46955
Qy 2398 actcggcgagcagatggattcttcaacgctcttagaataccatgatacccgctgccc 2457
Db 46954 ACTCGCGGAGCAGGATGATTCTTTCACACGCTTTAGGAATCCATGTATCCCGTGCCC 46895
Qy 2458 acgtcatta----- 2466
|||||
Db 46894 AGCTCATTAAGTTAGTGGACTGAATTTCTAGTGTATGCGCACACAATTCAAAACATAAT 46835
Qy 2467 -----agcaatcagagacatttgcgttttaaatccattcagtgctc 2508
|||||
Db 46834 GTTCTCTTTCCAATAATCAGCAATATCGAGAGCATTTGCTGTTTAAAGCCATTCAAGTGC 46775
Qy 2509 tggtagggcggggccacaataatgggagcgcttgcacacatcttctgtaagcgagagcg 2568
Db 46774 TGGTAGGAGCGGGCCACAAAATCGGAGGCGCTGCCACACATCTTCTGTCAGCGGAGAGCGG 46715
Qy 2569 aacagttagcaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2628
Db 46714 AACAGTTGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 46655
Qy 2629 ccctcttcaaggacactctctgggctgatgagatgagcgggatacaaacgttggtggt 2688
Db 46654 CCTCTTTAAGGACACTCTCTGGGCTGATGAGGATGACGGGATACAAACGTGCTGTGGT 46595
Qy 2689 gggcgatggtttggagaccgcagtcgactggctcttggtagaagacacgctggccgagc 2748
Db 46594 GGGCGGATGTTTGGAGACCGCACTGGCTGCTCTTGGTGAAGACACACGCTGGCCGAGC 46535
Qy 2749 aatttaccggcaggataaagcaaaatgccacgcagctcgaacagtcgagcggcaaaacgac 2808
Db 46534 AATTGACGGCAGGATCAAGCAAAATGCCACGACGCTCAACAGTGCAGCGCAAAACCTTC 46475
Qy 2809 atctgcccaagc-gctcatgctgtgctgagagct-aagatgatcttactaaaa 2861
|||||
Db 46474 ATCTCCCCAAGCAGCTGCATGCTGTGCTGCGAGCTCAAGATGATCTTACTACAA 46420
```

RESULT 5

AC009382/c
LOCUS
DEFINITION
Drosophila melanogaster clone RPC198-1B21, *** SEQUENCING IN
PROGRESS ***, 52 unordered pieces.
AC009382
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
1 (bases 1 to 92677)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Duan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
Tabo-P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Walbah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu.W., Zhou,X., Nelson,D. and
Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 92677)
Worley,K.C.
Direct Submission

COMMENT

AC009382 92677 bp DNA HTG 18-FEB-2000
Drosophila melanogaster clone RPC198-1B21, *** SEQUENCING IN
PROGRESS ***, 52 unordered pieces.
AC009382
GI:6996682
HTG: HTGS_PHASE1.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.
1 (bases 1 to 92677)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Duan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
Tabo-P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Walbah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu.W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 92677)
Worley,K.C.
Direct Submission
Submitted (20-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5881494.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRAU
Center clone name: RPC198-1B21
----- Summary Statistics

Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 42% of reads
Chemistry: Dye-terminator Big Dye: 24% of reads
Assembly program: Phrap: version 0.980611
Consensus quality: 38748 bases at least Q40
Consensus quality: 54037 bases at least Q30
Consensus quality: 64416 bases at least Q20
Estimated insert size: 135000; agarose-fp estimation
Estimated insert size: 83602; sum-of-contigs estimation
Quality coverage: 0.5x in Q20 bases; agarose-fp estimation
Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 828: contig of 828 bp in length
* 829 848: gap of unknown length
* 849 2184: contig of 1336 bp in length

FEATURES	source
1..57873	Location/Qualifiers
30893:	contig of 915 bp in length
30913:	gap of unknown length
30914	32447: contig of 1534 bp in length
30914	32467: gap of unknown length
32448	33518: contig of 1051 bp in length
32468	33538: gap of unknown length
33519	34583: contig of 1045 bp in length
33539	34603: gap of unknown length
34584	34604
34604	35755: gap of unknown length
35755	37161: contig of 1387 bp in length
35775	37181: gap of unknown length
37182	38633: contig of 1452 bp in length
38634	38653: gap of unknown length
38654	39940: contig of 1287 bp in length
39941	39960: gap of unknown length
39961	41986: contig of 2026 bp in length
41987	42006: gap of unknown length
42007	43675: contig of 1669 bp in length
43676	43695: gap of unknown length
43696	44832: contig of 1137 bp in length
44833	44852: gap of unknown length
44853	45381: contig of 1529 bp in length
45382	45401: gap of unknown length
46402	47470: contig of 1069 bp in length
47471	47490: gap of unknown length
47491	49244: contig of 1754 bp in length
49245	49264: gap of unknown length
49265	50791: contig of 1527 bp in length
50792	50811: gap of unknown length
50812	52579: contig of 1768 bp in length
52580	53599: gap of unknown length
52600	53959: contig of 1360 bp in length
53960	53979: gap of unknown length
53980	56252: contig of 2273 bp in length
56253	56272: gap of unknown length
56273	58783: contig of 1601 bp in length.

Query Match	13.9%	Score 524.4	DB 51	Length 57873
Best Local Similarity	97.6%	Pred. No. 3.3e-114		
Matches 531	Conservative 0	Mismatches 13	Indels 0	Gaps
Qy	170	agcattccgaatgcagacgacgacgatgaacttaatagacgctcgcgtggacgagtcacat	229	
Db	43132	AGCATTCCGAATGGACACGACACTGATGAACCTTATATACAGCTTCGCCTGGACGAGTCCAT	43191	
Qy	230	ggattgttcaaaagcggaggatgtcttcgaaccgttcgacccgactgcacctcgacatcggaacat	289	
Db	43192	GGATTGTGTTCAAAAGCGGAGGATGTCTTCGAACCGTTCGACGCGACCTGCACCTCGGACAT	43251	
Qy	290	gctggacatcatctctcaacgacatggacctggccgcgacgcagatgtgtacaacatgtcgtct	349	
Db	43252	GCTGGACATCATCTCTCAACGACATGGACCTGGCGCGACGCGAGATGTACAACATGCTGCT	43311	
Qy	350	ggacgagcctcgaacgcgatacccagcagaacgcagtcgcgtggatcagaagccgcaatccgt	409	
Db	43312	GGACGAGCCTCGAAGCGATACCCAGCAGCAGCTCCGTGGATCAGCAGCCGCAATCCGT	43371	
Qy	410	c9agcaacagccgcacgtgtaaaagcgagcactcttcgcgcagtgacatacgaaggaggaaact	469	
Db	43372	CGAGCAACAGCGGCACGTGAANAAGCGACACTCTTCGCCAGTGCACATCAAGGAGGAAC	43431	
Qy	470	gcatacgagcaacaacagtcgcgcgtcttcgtttacaacaacagatccccctcatagccac	529	
Db	43432	GCATCAGCAGCAACAACAGTCGCGCTTCTCGTCTACAAACACGATTCCTTCATAGCCAC	43491	

Qy	530	aagtacaattgttccccaagcaacgcgcagcgggcttttgaagccgcgcacacacagc	589
Db	43492	AAGGTACAAATTTGTCCTCCCAAGCAACAGCGACGCGGGCTTTTGAAGGGCGGCCCAACCAACAGC	43551
Qy	590	caccatcacatcacatgcagcgcacagcagatgcccccgaacacgcgcgtgttatcccccatc	649
Db	43552	CACCATACATCACATGACGCCCAAGCGGATGCCGCCGNAACAGCGGGTGTATTCCCCATC	43611
Qy	650	tctgggcagtagctttgtctactacagtcacatctccccgccacgcctgcgcgtggagctgc	709
Db	43612	TCTGGCAGTAGTCTTGTGCTCTACAGTCATCAGTCCATGNTCCCGCCACGTCGCGGGTGGAGTCTGC	43671
Qy	710	gaac	713
Db	43672	GAAC	43675

RESULT

7

AC009382

LOCUS

DEFINITION

Drosophila melanogaster clone RPI198-1B21, *** SEQUENCING IN PROGRESS ***, 52 unordered pieces.

AC009382

VERSION

AC009382.4

GI:6996682

KEYWORDS

HTG; HTGS_PHASE1.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

REFERENCE

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS

Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K., Bodota,B., Bock,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,B., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suggang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbeh,M., Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 92677)

AUTHORS

Worley,K.C.

TITLE

Direct Submission

JOURNAL

Submitted (20-AUG-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Feb 18, 2000 this sequence version replaced gi:5881494.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: DRAU
Center Clone name: RPI98-LB21

Summary Statistics
Sequencing vector: M13: I08821
Chemistry: Dye-primer Bodipy: 42% of reads
Chemistry: Dye-terminator Big Dye: 24% of reads
Assembly program: Phrap: version 0.980611

*	1	828:	contig of 828 bp in length
*	829	848:	gap of unknown length
*	849	2184:	contig of 1336 bp in length
*	2185	2204:	gap of unknown length
*	2205	3611:	contig of 1407 bp in length
*	3612	3631:	gap of unknown length
*	3632	4316:	contig of 685 bp in length
*	4317	4336:	gap of unknown length
*	4337	5322:	contig of 986 bp in length
*	5323	5342:	gap of unknown length
*	5343	6799:	contig of 1457 bp in length
*	6800	6819:	gap of unknown length
*	6820	7605:	contig of 786 bp in length
*	7606	7625:	gap of unknown length
*	7626	8704:	contig of 1079 bp in length
*	8705	8724:	gap of unknown length
*	8725	9498:	contig of 774 bp in length
*	9499	9518:	gap of unknown length
*	9519	10995:	contig of 1477 bp in length
*	10996	11015:	gap of unknown length
*	11016	11827:	contig of 812 bp in length
*	11828	11847:	gap of unknown length
*	11848	12755:	contig of 908 bp in length
*	12756	12775:	gap of unknown length
*	12776	13506:	contig of 731 bp in length
*	13507	13526:	gap of unknown length
*	13527	14778:	contig of 1252 bp in length
*	14779	14798:	gap of unknown length
*	14799	15619:	contig of 821 bp in length
*	15620	15639:	gap of unknown length
*	15640	17154:	contig of 1515 bp in length
*	17155	17174:	gap of unknown length
*	17175	18583:	contig of 1409 bp in length
*	18584	18603:	gap of unknown length
*	18604	19693:	contig of 1090 bp in length
*	19694	19713:	gap of unknown length
*	19714	21010:	contig of 1297 bp in length
*	21011	21030:	gap of unknown length
*	21031	22445:	contig of 1415 bp in length
*	22446	22465:	gap of unknown length
*	22466	23147:	contig of 682 bp in length
*	23148	23167:	gap of unknown length
*	23168	24468:	contig of 1301 bp in length
*	24469	24488:	gap of unknown length
*	24489	25609:	contig of 1121 bp in length
*	25610	25629:	gap of unknown length
*	25630	27332:	contig of 1703 bp in length
*	27333	27352:	gap of unknown length
*	27353	28358:	contig of 1006 bp in length
*	28359	28378:	gap of unknown length
*	28379	29948:	contig of 1570 bp in length
*	29949	29968:	gap of unknown length
*	29969	31673:	contig of 1705 bp in length
*	31674	31693:	gap of unknown length
*	31694	33056:	contig of 1363 bp in length
*	33057	33076:	gap of unknown length
*	33077	34658:	contig of 1582 bp in length

[illegible]

gccaagaccctggagaaggctgggcgacaaagctcaaggagtgtctaccagctgat


```
QY 1901 ggcgttgaggttacttgcagaccagctggcaggttcccttcgtttctttccatcgcc 1960
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2023 CCACCTGGACCTGGCTTGAGCCTCTCTGGAACCTCATCCGTCACCTGCTGAGCGTCT 2082
QY 1961 ctggctgggctggggtgctgtcaaggcggtccggtgggctgttttagcaaacgcgccagcag 2020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2083 CTGGGTGGCGCGCTGCTGGCAGCGCGGCGAGGGGCGCTGGCAGCAGTGCTCTGCG 2142
QY 2021 gaaacaggcgctggaatctgtcacgcgaactggccctgcctcttcaaacgactgaataatt 2080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2143 AGTGGATGTAGCGCCAGCGCCCGCAGCAGCGCCCTGGCTTACATTAAGCTGCACCACT 2202
QY 2081 gcaactgactggaatggaagcgcggtgacatgaacggtattatgatgacactattcgc 2140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2203 GCACACATGGGAAGACACACAGCGGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262
QY 2141 aagcaaatggtggaagtggcgcaaatctactgacacgcgcgagaccatctgcaccca 2200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2263 CCTGAACCTGGCAGAGTGTGCAGGGATGCCGTGTCTGTGGCAGCCTGGCCGAGATCTA 2322
QY 2201 cgttaatgaacgcttgcgaatgaagcgcagtgccccaataatggttgcaacagttctcgc 2260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2323 TGTGGCGGCTGCAATGAGAGTGAAGACCAAGTCTCCACCGGCCCTTG---CATTTTCTGAC 2379
QY 2261 ccgatactacatgagccgggctcgtcaagagtcggtcgcactagggccaccgcagcaaac 2320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2380 ACGCTTCTTCTGAGCAGTGCCTCCCGCAGGCTCGCTGGCMACAGAGTGGTCACTGTC--- 2436
QY 2321 cgaggagctacgttgggcatcagcctatgatatcgtactgcccacgcacgctctt 2380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2437 TCCTGCCATGCACTGGCTCTGCCACCCGCTGGGCCACCGTTTCTCG---TGGATGGGA 2493
QY 2381 caogtcagatctgagcgaactccggcagcaggatgatttcttcacagctcttaggaatcc 2440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2494 CTGTCCTGCTCAGTACCCCATG-----GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547
QY 2441 atgtgatccgctgcccacgctgaattaaagcaatcagagcatttgcgtttaataccat 2500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2548 AGTGGACCCCTCGCCCGCAGTGACTCAGCTATTCCGGGAACATCTTTAGAGCCGACACT 2607
QY 2501 tcagtctgtgtaggagcgggccacaaatcgggagcgcttgcacacatcttcttgcagcgg 2560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2608 GAACTGTGTACCCAGCCCAACCCACGCCCTGGGTGAGTGGGACAAAGCAATTCCTC 2667
QY 2561 agaggcggaacagttgcagcaaacagcagcacagcggcaccattgtcagaaattcttaa 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2668 GGATGCCCTCGGTACTGCAGCTGTGTGAACAGCTGTTCTGTGCTCGCGGGGCTCTCTGC 2727
QY 2621 gtacagctccctcttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2728 CTACAGCTTCTCATCAGTTCCA---GCATGGCCACCACCACCGCGGTAGACCCGGTGGC 2784
QY 2681 ggtggtggggcgatgttttgagaccgcagtgactggctcttggtagaacacgcgt 2740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2785 CAAAGTGGTGGGCTCTCTGACAGCTGTGTGTATCCACTGGCTGCGCGGGGATGAGGAGC 2844
QY 2741 ggcgcggaattgtcagcgaggatacagaataatgcccacgcagctgcaacagtgcgcgga 2800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2845 GGTGTAGCGGCTGTGCCCGCTGTGTGAGCACCTTGCCCGGCGGTCTGCAGGAGTCTGAGAG 2904
QY 2801 aaacgatcatctcccaaggcgtgcatgctgtgctgagcgaataatgatcttactaaa 2860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2905 ACCCTTCCCGAGGC---AGCTCTGCATCTCTCAAGGCTGCCCGGCCCTGCTGGGCTG 2961
QY 2861 aaacaatggcaacgcactggacaaaagtctcaagcaattgttaaacatctcttgcgatga 2920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2962 TGCAR--GGCAGAGTCTGGTCCAGCCAGCCTGACCATCTGTGAGAAGGCCAGTGGGTACC 3020
QY 2921 gtcagtgtygagctccaagagtgtcttgcgtgtcaaccggatcacccgaocgaagggat 2980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3021 TGCAGGACAGCCTTACCCACACCAG--CCAGCAGCTCCATTGCAAGGCCCTGTCAGCT 3078
```

```
QY 2981 aaagctgttttccagttgcttacctgcgattggctgctgcgaactaggactgctctgtg 3040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3079 GTTCTCTGTGACCTGCTTCTTGTGGTGCACACGAGCTTGGCGGAGCAGAGCACCCCC 3138
QY 3041 ggaactggaacacatgaatatggaggacgatggtcttaccacagtgccaggtgaagtgtct 3100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3139 GGCCTCCGGCCAGCAGCAGCCAGCGCGCCAGCAGCAGGCCCCAGGCTTCCGCCCTTGAGCT 3198
QY 3101 cgaagaagttccgaacgcgatttgaactcgttgcgaacatgtggagaatataccgaaacgc 3160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3199 GCCTGGCTTCCAACGGGACCTTGAGCAGCTGAGCGGCTGGCAGAGCTTCCGGCCCGC 3258
QY 3161 ccaatcgccatataattgtacgagcgagtttgtcgctctgatggctgagcctcacctgtg 3220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3259 CATCGCGAGGCTGTTCTTACATGAGGCCACGCCCGCTGATGGCGGGGCCAGCCCCAC 3318
QY 3221 tccaacgcgaacagcttctggacagggagtctgcatacgcgaacgccacactcttcattt 3280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3319 ACGGACACACAGCTCTCGACCGCAGCTTGAGCGCGCGGCGAGCCCCCGTGGCAAGG 3378
QY 3281 ctggcgagcaagatcggcgagcagaacttcgtggcgagagagcggaaacgggcttc 3340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3379 AGCGCGGCTGGCGAGCTGGAGCCGC-----GGCCACGCGCGGAGCACGCGGA 3429
QY 3341 ggcctgtacgtggcctgcaagtatctccgcctgcgctgcctcagctcccggtgaagc 3400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3430 GGCTTGTCTCTGGCTTCTTACCTGCCCCCGGCTTCTGTGCGCGCCCGGGCAGCG 3489
QY 3401 tgcctgcatgttagcgcgagcgcccaagacccttggagaagtggtggcgacaagcgaagct 3460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3490 CGTGGGCATGCTGCTGAGCGCGCGCACACTCGAGAAGCTTGGCGATCGCGGCTGCT 3549
QY 3461 caaggagtcctaccagctgatgaagtcgtggtggcaacgcga 3501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3550 GCACGACTGTACAGATGCTCATGCGCCTGGCGGTGGGA 3590
```

RESULT 9

AR069298 LOCUS AR069298 4154 bp DNA PAT 18-FEB-2000

DEFINITION Sequence 37 from patent US 5891631.

ACCESSION AR069298

VERSION AR069298.1 GI:7220186

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 4154)

AUTHORS Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.

TITLE Methods relating tosterol regulatory element binding proteins

JOURNAL Patent: US 5891631-A 37 06-APR-1999;

FEATURES

Location/Qualifiers

source l..4154

BASE COUNT 714 a 1406 c 1280 g 754 t

ORIGIN

Query Match 4.4%; Score 165; DB 5: Length 4154;
Best Local Similarity 45.7%; Pred. No. 1.4e-28;
Matches 1144; Conservative 0; Mismatches 1295; Indels 62; Gaps 14;

```
QY 1023 gaagtaagcgtcggcccaacgcacatcgagcgcgctatcgacatcaatacaacgac 1082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1130 GGAGAGAAGCGCACAGCCCAACGCCCAATTGAGAAGCGCTACCGCTCTCCATCAATGAC 1189
QY 1083 aagattaacagttgaagaacttgtagtggagagcagcccaagctgaacaagtcgcga 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1190 AAAATCACTTACGACTCAAGGATCTGTGGTGGGCACTGAGGCAAAAGCTGATAAATCTGCT 1249
QY 1143 gttgtcggaatcccatagacaagattcgggattctgcaacgccagaatcacgattcgaag 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1250 GTCTTGGCAAGGCCATCGACTACATTGCTTCTTGTCAACACAGCAACCACTCAAG 1309
```

QY 1203 gcagagttgcagcgcctgcagaggaagctaatggcagcgcagcgcctccaaggtgaagat 1262
DB 1310 CAGGAGAACCTAAGCTGCGCACTGCTGTCCACAAAAGCAAACTCTGAAGGATCTGGTG 1369
QY 1263 t----tacttcagctggcactcggcctggttagagcaatccaagaagcgcgcgagagctc 1318
DB 1370 TCGGCTGTGGCAGTGGAGGGAACACAGACGTGCTCATGGAGGCGTGAAGACTGAGGTG 1429
QY 1319 gcagacctttaccacaggtatgccggactgcagcgcgcagcgcagcagatcgcgataccttc 1378
DB 1430 GAGGACACACTGACCCACCCCTCGGATGCTGGCTTCACCTTTCCAGACAGCAGCCCTTG 1489
QY 1379 gctcgcgccatgcactcgcacatctctgttgcgcgcatacccttatggtgatccaccgc 1438
DB 1490 TCCCTGGCAGCAGGGCAGTGGCAGCGGTGGCAGTGGCAGTGCAGTGCAGACCTGCACAG 1549
QY 1439 cagctgtagcagtggcagcgcagcagcaatgaagaaccactggtgcccagctctat 1498
DB 1550 CAGTCTTTGAG-GACAGCAAGGCAAGCCAGAGCAGCGGCCGTCTCTGCACAGC----- 1603
QY 1499 gcgcgcgatggcaccacactcctgcctcgcagctcgtcatgtttatgttgcacctcctggc 1558
DB 1604 -CGGGGATGCTGGACCGCTCCCGCCTGGCCCTGTGCACGCTGCTCTTCCTGCTGTC 1662
QY 1559 cgtcaatcccttcaagacctttctccagcgcgcacactatgacagtaatgacgatctgg 1618
DB 1663 CTGCAACCCCTTGCCCTCTTGCTGGGGGCCGGGGGCTTCCACAGCCCTTCAGATACCAC 1722
QY 1619 cgcacatgagc-----ggtcaagacgcattctctcttaccagctggaaggtgaag 1669
DB 1723 CAGGGTCTACCATAGCCCTGGCGCAACGTGCTGGGCACCCAGAGCAGAGATGGCCCTGG 1782
QY 1670 ttttgcgtcgcagcagagttcctcgttatgagcttatggaacttcaactcagctgatctgg 1729
DB 1783 CTGGGCCACGTGCTGCTGCCCCAGTGGTCTGGCTCTCAATGGGGTGTGTGGTCTGCT 1842
QY 1730 atgcttggtgaattgctgtttacagtgatcc-----gcagctggcagcgcgaac 1780
DB 1843 CTCTTGGTCTCTCTTTGTGTACGGTGAGCCAGTGCACAGGCCCTCAGTGCAGCCCGC 1902
QY 1781 ggaocctaetgcagcacagcagcggcgtgacttcttatcccaagagacagtgctc 1840
DB 1903 CGTGTACTTCTGGAGGCATCGCAAGCAGGCTGACCTGGACCTGGCCGGGGAGACTTTGC 1962
QY 1841 tcagggctacgcgggttaacctcaactgtgcatagtttggattaaagtatacgcggctc 1900
DB 1963 CCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGGCGGCACTGGGGCGGCCCTGGCCACCTC 2022
QY 1901 gcgcttgagtggttaetgcagaccacgtggcagttcctctctgttttcttccatcgct 1960
DB 2023 CCACCTGGACCTGGCTGTAGCCTCTCTGAACTCTATCCGTACCTGCTGCAGCGTCT 2082
QY 1961 ctggtggtggtggtgctgcagcggcgttcggtggtggtggtttagcaacccgcagcag 2020
DB 2083 CTGGTGGGCGCTGGCTGGCAGCGCGGCGAGGGGCTGCAGCAGGACTGTGCTTCGC 2142
QY 2021 gaaacagcgtggcatttcgcagcgaactggccctcgtcttccaccgcagctgaataatt 2080
DB 2143 AGTGGATGCTAGGCGCCAGCGCCGAGCAGCGCCCTGGCTTACCATAAAGTGCACCACT 2202
QY 2081 gcaactgactggaattggaagcgcggtgacatgaacgcgcattatgatggcactattcgc 2140
DB 2203 GCACACCATGGGAGACACAGCGGGGCACTTCACTGCCACCACTGGCGCTGAGTGC 2262
QY 2141 aagcaacatggctgaagtggcgacaaatctactgacacccgcgcgagaccatctgcacaa 2200
DB 2263 CCTGAACCTGGCAGAGTGTGCAGGGGATGCCGTCTGTGGCAGCAGCTGGCCGAGATCTA 2322
QY 2201 cgttaatgacagcgttgcgaatgaagcgcagtgccccaaaatggttgcacagttcttcgc 2260
DB 2323 TGTGGCGGCTGCAATTGAGAGTGAAGACCACTGCTCCCAAGCGGCGCTTG---CATTTCTGAC 2379

QY 2261 ccgatactacatgagccgggctcgtcaagagtgcggtcgcactagggccaccgcagcaaac 2320
DB 2380 ACCTTCTTCTGAGCAGTGCCTGCCAGGCTGCTGGCACAGATGGCTCAGTGC---C 2436
QY 2321 gcaggagctacgttgggcatcaccagcctatcgatactcgtactgogccacgcagctctt 2380
DB 2437 TCTGTGCATGAGTGGCTGTCACCCCGTGGCCACCGCTTCTTCTTCG---TGGATGGGA 2493
QY 2381 cagctacgatctgagagactcccgagcagagatggtattcttcacagctcttaggaatcc 2440
DB 2494 CTGGTCGGTCTCAGTACCCCATG-----GGAGAGCTGTACAGCTTGGCCGGGAACCC 2547
QY 2441 atgtgacccgcctgcaccacgtcattaaagcaatatcgagagcatttgcgttttaaatccat 2500
DB 2548 AGTGGACCCCTGGCCAGGTGACTCAGCTATTCCCGGAACATCTCTTAGAGCAGGACT 2607
QY 2501 tcagtgctcgttggagcggggcccaaatcgggagggcctgcccacatctcttctcagcgg 2560
DB 2608 GAACTGTGTACCCAGCCCAACCCAGCCCTGGGTACGCTGATGGGCAACAAGGAATTCTC 2667
QY 2561 agagggcgaacagttgcagcaacagcagcagcagcggcaccattgtcagcaatgttctaa 2620
DB 2668 GGATGCCCTCGGTACCTGCAGCTGTCTGAACAGCTGTCTGATGCTGCGGGGCTCTCGC 2727
QY 2621 gtacaagtcctccttaagacacactctcgtggtgatgagatgagcgggatacaaacgt 2680
DB 2728 CTACAGCTTCTCCATCAGTTCCA---GCATGGCCACACACCCGCGGTAGACCCGTTGC 2784
QY 2681 ggtgtggtggcgcagtgttttgagaccgcagtgcactggtcctcttggtaagacacgct 2740
DB 2785 CAAGTGTGGGCTCTCTGACAGCTGTGGTGATCCACTGGCTGCGCGGGATGAGGAGGC 2844
QY 2741 ggcgagcaattgacgcgaggaatcaagcaaatgccacgcagctgcaacagtcgcgcga 2800
DB 2845 GGTGAGCGGCTGTCGCCGTGGTGAGACACCTGCCCGGGTGTCTGAGGAGCTGAGAG 2904
QY 2801 aaacgatcactcgcacaaagcgtgcattgctgctgcgagctaaagtatcttactaaa 2860
DB 2905 ACCCTGCCCAGGCG---AGCTCTGCATCTCTTCAAGGCTGCCCGGCCCTGCTGGGCTG 2961
QY 2861 aaacaaatggcaacgcactggacaaaaagctcgaagcaattggtaaaaatcctctgcgata 2920
DB 2962 TGCCAA--GGCAGAGTCTGTGTCACGCCAGCTGACCATCTGTGAGAAAGCCAGTGGGTACC 3020
QY 2921 gtcgagtggtgagctcgaagtgcttgactgtcaaccggtatcaccgacgcccaaggtat 2980
DB 3021 TGAGGACACCTTGGCTTACCACACCAG--CCAGCAGCTCCATTGACAAAGCCGCTGCAGCT 3078
QY 2981 aaagctgttttccagttgcttacctgcgattgctgctcgaactaggactgctctgtg 3040
DB 3079 GTTCTGTGTGACCTGCTTCTTGTGGTGCACACAGCTGTGGCGGCAGCAGCAGCCCC 3138
QY 3041 ggaactggaacacatgaatatggagagcagtggtcttaccagtgccaggtgaagtgcct 3100
DB 3139 GGCCTCGGCCCCAGCAGACCCAGGCGCACACAGCAGCGCCCGCCAGGCTTCCGCCCTTGA 3198
QY 3101 cgagaagttccagacccgatttgaactgttcgcaaacattgtggaataataaccgaacgc 3160
DB 3199 GCCTGGCTTCCACAGGACCTGAGCAGCTGAGCGGCTGGCAGAGCTTCCGGCCCCG 3258
QY 3161 ccaatgcgcgcatatatgtttacgagggcagttgtcgcgtgatggtgagcctcaccgtg 3220
DB 3259 CATGCGGAGGGTGTCTCTACATGAGGCCACGGCCCGGCTGATGGCGGGGCCAGCCCCAC 3318
QY 3221 tcaacgcgaacagcttggagcagggagctgcgatacagcaacgcgccactcgtccactctt 3280
DB 3319 ACAGACACACAGCTCTCTGACCGCAGTCTGAGCGCGGGGAGGCGCCCGGTGGCAAGG 3378
QY 3281 ctgcgcgacgaagatcgcgcgcagcagaactcgtggcggagagcgggaacgggcttc 3340
DB 3379 AGCGCGGCTGGGAGCTGGAGCGCG-----GCCACCGCGGGGAGCAGCAGCGGA 3429
QY 3341 ggccatgtacgtggcctgcaagtatatctccgcctgagctcagctcccccggtgaacg 3400

Db 3430 GGCTTTGCTGCTGGCTCTGCTACCTGCGCCCGCGCTCTCTGTCGGCGCCCGGGCAGCG 3489
Qy 3401 tctgtgcattgttagcgcagcggcccaagaacctctggagaaggctgggcagcaacgcaagct 3460
Db 3490 CGTGGGCATGCTGGCTGAGCGCGCGCACACTCGAGAAAGCTTGGCGATCGCGGCTGCT 3549
Qy 3461 caagagtgctaccagctgatgaagtcgctgggcaacgcga 3501
Db 3550 GCACGACTGTCAGCAGATGCTCATGGCCCTGGCGGTGGGA 3590

RESULT 10
LOCUS I22310 4154 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 37 from patent US 5527690.
ACCESSION I22310
VERSION I22310.1 GI:1602664
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4154)
AUTHORS Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
TITLE Methods and compositions relating to sterol regulatory element binding proteins
JOURNAL Patent: US 5527690-A 37 18-JUN-1996;
FEATURES Location/Qualifiers
source 1..4154
/organism="unknown"
BASE COUNT 714 a 1406 c 1280 g 754 t
ORIGIN

Query Match 4.4%; Score 165; DB 5; Length 4154;
Best Local Similarity 45.7%; Pred. No. 1.4e-28;
Matches 1144; Conservative 0; Mismatches 129; Indels 62; Gaps 14;

Qy 1023 gaaglaaagcgtcgcccaacgccatcgagcgcgctatcgcaacctcaatcaacgac 1082
Db 1130 GGAGAGAACGGACACGCCACAAACGCCATTGAGAGCGTACCGCTCTCCATCAATGAC 1189
Qy 1083 aaagattaacgagttgaagaacttggtagtgagagcagcgcccaagctgaacagtcgcga 1142
Db 1190 AAAATCATTTGAGCTCAAGGATCTGGTGGTGGCACATGAGGCAAAAGCTGAATAAATCGCT 1249
Qy 1143 gtgttcgcaaatccatagacaagattcgggattctgcacgcgcagcaatcacgattcgaag 1202
Db 1250 GTCTTGGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAAGAACTCAAG 1309
Qy 1203 gcagagttcagcgcctgcagaggagactaatggcacgcagcgcgtcccaagtggaagat 1262
Db 1310 CAGGAGAACTTAAGTCTGCGCACTGTGTGCCAAMAAGCAAAATCTCTGAAGGATCTGGTG 1369
Qy 1263 t----tacttcagctgggcactcggcctggtagagactccaagaagcgcgcgcgagagctc 1318
Db 1370 TCGGCCGTGGCAGTGGAGGAAACACAGAGCTGCTCATGGAGGGCGTGAAGACTGAGGTC 1429
Qy 1319 gaagacctttacaagatgcccgcgactgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1378
Db 1430 GAGGACACACTGACCCCAACCCCTCGGATGCTGGTTCACCTTTCCAGAGCAGCACCCCTTG 1489
Qy 1379 gctctgcgcctgcactcggacatctcgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1438
Db 1490 TCCCTTGGCAGGAGGGCAGTGGCAGCGGTGGCAGTGAGGAGTGACTCGGAGCGTGCACAGC 1549
Qy 1439 cagctgtagcagtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1498
Db 1550 CCAGTCTTTTGGAG-GACAGCAAGCAAGCCAGAGCAGAGCGGCCGCTCTCTGCACAGC----- 1603
Qy 1499 ggcgcgcagtgccaccccaactcgcctcgcgaactctgcattgttatgttgcacatcctgcgc 1558
Db 1604 -CGGGGCATGCTGGACCGCTCGCGCTGGCCCTGTGCACGCTCGTCTTCTCTGCGGCTGTC 1662

Qy 1559 cgtcaateccctcaagaccttttctccagcgcgcgcactatgacaglaaatgacatcttgg 1618
Db 1663 CTGCACACCCCTTGGCCCTCCTTGTCTGGGGGCCCGGGGCTTCCCAGCCCTCTCAATACCAC 1722
Qy 1619 cgacatgagc-----ggtcaaaagacgcatcttctcttaccgacgtgaaagtggaag 1669
Db 1723 CAGCGTCTACCATAGCCCTGGGCGCAACGCTGGCACCAGAGAGCAGACATGGCCCTGG 1782
Qy 1670 ttttgctgtctggcagcagagttccttgatatgcttattgaacttcaacactgatgttgg 1729
Db 1783 CTGGGCCAGTGGCTGCTGCCCCAGTGGTCTGGCTGCTCAATGGGCTGTGTGGTGTCTGT 1842
Qy 1730 atgcttggtgaaattgctggtttacggtgatcc-----gcagctggagcgcgcaaac 1780
Db 1843 CTCTCTTGGTGTCTTCTTGTCTACGGTGAGCCAGTCACAGCGGCCCTCACTCAGGCCCCG 1902
Qy 1781 ggacgcctactgccagcacagcggcggcgtgacttcttattttagccaaagacagtcgtc 1840
Db 1903 CGTGTACTTCTGGAGGCATCGAAAGCAGGCTGACCTGGACCTGGCCGGGGAGACTTTGC 1962
Qy 1841 tcaggcctacgcgcggttaacctcaactgctgcataatgtttggattaaagttaacgcgcgtc 1900
Db 1963 CCAGGCTGGCCACAGCAGCTGTGGCTGGCCCTTGGCGGCACATGGCGCGGCCCTGCCACCTC 2022
Qy 1901 ggcgttgagtggttaacttgcagaccacgctggcagttcctcttcttcttccatcgcct 1960
Db 2023 CCACCTTGGACCTGGCTTGTAGCCTCTCTGGAACCTCATCCGTCACCTGCTGACGCTCT 2082
Qy 1961 ctggtcggttcgggtgctgtcacgcgggtccggtcggtgttttagcaacgcccccagcag 2020
Db 2083 CTGGTGGCGCGCTGGCTGGCAGCGCGGACGGGGCTGCACGAGGACTGTCTCTGCG 2142
Qy 2021 gaaacaggcgcgtgctatctgcacgcgaactggccctgcgtcttcaacgcgactgaatacatt 2080
Db 2143 AGTGGATGTAGCGCAGCCGCCAGACGACGACCCCTGGTCTACATAAGCTGCACCACT 2202
Qy 2081 gcaactgactgaaaaatggagcgcggtgacatgaacgcgcatattgatggcactattcgc 2140
Db 2203 GCACACATGGGAAAGACACACAGCGCGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262
Qy 2141 aagcaacatggctgaagtggcgcacaaatctactgacacgcgcgcgagaccatctgcacca 2200
Db 2263 CCTGAACCTGGCAGAGTGTGAGGGATGCCGTGCTGTGGCGACGCTGGCCGAGATCTA 2322
Qy 2201 cgttaatgacagcgttcgcgaatgaagcgcagtgcccccaaaaatggttgcaacagttctcgc 2260
Db 2323 TGTGGCGGCTGCATTGAGAGTGAAGACCACTCTCCACGGGGCTTG---CATTTTCTGAC 2379
Qy 2261 ccgatactacatgagcgcggcctcgtcaagtgctcgcgcactagggcccccgcgagcaaac 2320
Db 2380 ACGCTTCTTCTGAGCAGTGCCTGCCCGCCAGGCTGCTGGCACACAGTGGCTCAGTGC---C 2436
Qy 2321 cgaggagctacgttgggcattcacagcctatggatatcgcctactgcgcacgcacgctctt 2380
Db 2437 TCCTGCCATGCAGCTGGCTCTGCCACCCCGTGGCCACCCGTTTCTTCG---TGGATGGGA 2493
Qy 2381 caagtcagatctgagcgcactccgcgcagcaggaatgagttcttcaacgcttcttaggaatcc 2440
Db 2494 CTGTGCGGTGCTCAGTACCCCATG-----GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547
Qy 2441 algtgtcccgctgcccacgctcattaaagcaatcatcagagcatttgcgtttaaatccat 2500
Db 2548 AGTGGACCCCTTGGCCCAAGTGACTCAGCTATTCCGGGAACATCTCTTAGAGCGAGCACT 2607
Qy 2501 tcagtgctctggtaggcgcgcgcacaaaatcgggagcgctgcccacatcttctgtcagcgg 2560
Db 2608 GAACCTGTGACCCAGCCCAACCCCAAGCCCTGGTGGTTCAGCTGATGGGACAAGAAATCTC 2667
Qy 2561 agaggcggaaacagttgcagcaacagcagcagcagcagcagcagcagcagcagcagcagcagc 2620
Db 2668 GGATGCCCTCGGGTAGCTGCAGCTGCTGAAACAGCTGTTCTGTATGCTGCGGGGCTCTCTGC 2727

QY	2621	gtacagctccctcttaagagacactctctgggctgatgaggatgagcgggatacaaacgt	2680
Db	2728	CTACGCTTCTCCATCAGTTCCTCA--GCATGGCCACACACCGCGTAGACCCGGTGGC	2784
QY	2681	ggtggtggtggcgatggtttggagaccgcagtgactgctcttgggtgaagacgct	2740
Db	2785	CAAGTGGTGGGCTCTCTGACAGCTGTGTGATCCACTGGCTCGCGGGATGAGGAGC	2844
QY	2741	ggccgagcaattgtacgagcagatcaagcaaatgccacgcagcagcagcagcagcga	2800
Db	2845	GGCTGAGCGGCTGCGCGTGTGGAGACACCTGCCCGGGTCTGCAGGAGTCTGAGAG	2904
QY	2801	aaacgatcatctgcccaagcgcgtgatgctgtgctgcgagctaaagtatcttactaa	2860
Db	2905	ACCCCTGCCAGGCG--AGCTTGCACCTCTTCAAGGCTGCCCGGGCTGTGGGCTG	2961
QY	2861	aaacaatggcaacgcactggcaaaaagttctcaagcaatggttaaacatctctgcgatga	2920
Db	2962	TGCCAA--GGCAGAGTGTGTCTCCAGCCAGCCTGACCATCTGTGAGAGGCGAGTGGTACC	3020
QY	2921	gtcgaagtgtgagctccaagagtgttactgtcactgtcaaccggatcaccgacgcgaaggtat	2980
Db	3021	TGCAGACAGCCTGGTACCACACACAG--CCACAGCTCCATTGACAAAGCCCGTGCAGCT	3078
QY	2981	aaagctgttttccagttgcttaccctgcgattggtgctgcgaactaggaactgctctgtg	3040
Db	3079	GTTCCTGTGTGACCTGCTTCTTGTGTGGCGACACAGCCTGTGGCGCAGCAGCAGCCCC	3138
QY	3041	ggaactggaaacatgaatatgagagacgatggtctctaccgaagtccagagtgct	3100
Db	3139	GGCCCCGGCCCCAGCAGCCAGGCGACACAGCAGCAGCGCCCCAGGCTTCCGCCCTTGCAGCT	3198
QY	3101	cgagaagtccagaccgattgaactcgttgctgcgaacattgtggagaataaccgaacgc	3160
Db	3199	CGTGGCTTCCACGGGACCTGACAGCCTGAGCGGCTGAGCGGCTGGCACAGACTTCGGGCCCCG	3258
QY	3161	ccaatgcgcgcatattgttacgagcagttgttcgctgctgatgctggagcctcaacgtg	3220
Db	3259	CATCGGAGGGTGTCTCTACATCAGGCGCACGCGCGCGCTGATGGCGGGGCCAGCCCCAC	3318
QY	3221	tcaacgcgaacagctcttgaaaggagtctgcgatcacgcgaacgccactcgtccatctt	3280
Db	3319	ACGGACACACAGCTCTCTCAGCCGAGCTGTGAGCGCGCGGCGAGCGCCCGGTGGCAAAGG	3378
QY	3281	ctgcggcagcaagatctgcgcgcgagcagaactctgtggcgcgagagcgggaacggcttc	3340
Db	3379	AGCGCGGTGGCGAGCTGGAGCGCG-----GCCCCACCGCGGGAGCACGCGGA	3429
QY	3341	ggccatgtacgtggttcgaagtatatctccgcctgcgtgcgtcagctccccgggtgaacg	3400
Db	3430	GGCCTTGTCTGCTGCTTCTTCTACCTACCTGCCCGCGGCTTCTCTGCGCGCGCGGCGAGCG	3489
QY	3401	tgctgcagttagccgagcgcgcaagaccctggagaagtggtggcgcaagcgaaagct	3460
Db	3490	CGTGGGCATCTGCTGGTGGCGCGGCGGCACACTCGAAGACTTGGCGATCGCCCGCTGCT	3549
QY	3461	caaggagtgctacagctgatgaagtgcgtgggccaacgcga	3501
Db	3550	GCAGGACTGTACAGAGATGCTCATGCGCCTGGCGGTGGGA	3590
RESULT	11		
HSU02031			
LOCUS	4249 bp	mRNA	PRI 22-OCT-1994
DEFINITION	Human sterol regulatory element binding protein-2 mRNA, complete cds		
ACCESSION	U02031		
VERSION	U02031.1	Gr:451329	
KEYWORDS	SREBP-2.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		

REFERENCE	1	(bases 1 to 4249)	
AUTHORS	Hua,X., Yokoyama,C., Wu,J., Briggs,M.R., Brown,M.S., Goldstein,J.L. and Wang,X.		
TITLE	SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (24), 11603-11607 (1993)		
MEDLINE	94089681		
REFERENCE	2	(bases 1 to 4249)	
AUTHORS	Hua,X.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-1993) Xianxin Hua, Department of Molecular Genetics, University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235 USA		
FEATURES	Location/Qualifiers		
Source	1..4249		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	118..3543		
	/standard_name="SREBP-2"		
	/codon_start=1		
	/product="sterol regulatory element binding protein-2"		
	/protein_id="AAA50746.1"		
	/db_xref="GI:51330"		
	/translation="MDDSGELGGLETMETLDELGLDIDEMLQFVSNVGVGFEPD LFSEQLCSFGSGSGSGSSSSNGSGSGSAGVDFSPQSTQVTLPSFSP SAASPQAPTQVKVSTVPTPRATPILQPRQPQPOQPOQLOQQTVMITFTSTP QTRIIQPLIYONAAATSFVLQPOVSLVTSOVOPVTIQOQOVQAOQVLTQTANG TLTQTAPATVOTVAPOVOQVPLVPOQIITKDTLSLTLTKTDGSPVMAVONPALTA LTPTQTALQVTLVSGSGTTLTPVMGQEKVPIKQVPGCVKQLEPKPKGERRTT HNIIBKRYSSINDKIIELDLVMDGTDAKMHKSGVLRKAIDIKYIQQVNHKLROENM VLKANQKLLGIDLGSLVDNEVDLKIEDFNQVLLMSPADSGSQAGSPSYID SEPGPLLDADKDEPPPPVALGVDRSLTLLVLCVLPCLISFNPLTSLQWGAHD SDQHPHSGSGSVLSESGSGGFWDMPTLLLLVNGVIVLSVFVKLIVHGEVTRP HRSRYTPWRHRKQADLDLARGDFAAANAQLQCLAVLGRALPTSRDLACLSSMNI RYLSQKRLVRLWLAKVFCRRATPATEGFEDEATSRDAAALYHRLHQHQTGKL PAGESCDVHMALCANVLAECAEKIIPSTLVEIHLTAAGLKTGCGKGLFLASTFL SRAQSLCGPHEASVPDSLRLWLCPLGKFMEFWSVSKAESLYCAORNPAIDIAQ VHQAFCKNLERATISLVKPAQKAGQOEESCFSSALEYKLLHSFVGVMSYPLSRSSVLKSAUGPDIICRWMTSAITVAISWLOGDDAAVRSHTKVERIPKALEVTES PLVKAIFHACRAMHASLPGKADQOSSCHCERASGHLWSLNSVSGTSDPALNHVQY LLTCDLLSLRTALWQKASASQAVGETYHASGABELAGPQDLGSLRLHLSHFRPAYR KHLFHEATVRLMAGSGPTHTHOLLEHSURRTTQTKHGEVDAMWQQRERATILLAC RHLPLSFLLSSPGQRAVLAAEAATLEKVGDRRSCNDCCQMIVKLGGGTAIAAS"		
BASE COUNT	860 a	1315 c	1202 g
ORIGIN	872 t		
Query Match	3.0%	Score 114.4;	DB 9; Length 4249;
Best Local Similarity	52.3%	Pred. No. 1.7e-16;	
Matches	315;	Conservative 0;	Mismatches 266; Indels 21; Gaps 2;
QY	2909	cctctgcgatgagtcgagtgtagctccaagagtgcttgactgctcaacgcgagatcaccca	2968
Db	2946	CCATTCCGAGAGGCCAGTGCGCCACCTATGAGACACGCTCAAGCTCAGTGGGGGCACCTC	3005
QY	2969	cgccaagggtataaagctgtttccagtgcttacctcgattgctgctgcgaactag	3028
Db	3006	TGACCCTGCCCTCAACACACCTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTGCTACG	3065
QY	3029	gactgctgtggaactggaaacacatgaatatggaggacgatgg--cttctacaagt	3085
Db	3066	GACAGCGCTCTGGCAAAAACAGGCCAGTCCAGCCAGGCTGTGGGGGAGACCTTACCACGC	3125
QY	3086	gccagtggaagtgcgcgaagttccagaccgattggaactcgttcgcaaacatttgga	3145
Db	3126	GTGAGCGCTGAACCTGCGGGCTTCCACGGGACCTTGGCAGGCTCGGAGGCTGSCACA	3185
QY	3146	gaataaccgaacgcccaatcgcgcataattttgtacgagggcaggtttgtgcctgatggc	3205
Db	3186	CAGCTTCCGCGCCAGCATACCGCAAGGTGTCTCTGCATGAAGCCACCGCTGCGCTGTGCG	3245
QY	3206	tggagcctcacctgtgccacgcgaacagctcttggacagagtgctgcgatcacgcaacgc	3265

Db	1804	CAGCTGTGGCTGGCCCTGCGAGCATTTGGGACAGGCCCTGGCCCACTCGAACCTGAGACTTG	1863
Qy	1914	tacttgcagaccacgtggcagttctctctgattttcttccatgcctctgcgtgggtcg	1973
Db	1864	GCCTGCAGCCTGCTTTTGAACCTTCATCCGCCACCTGCTGCGAGCTCTCTGGTTGGCCG	1923
Qy	1974	gtcgtgtcacgcgcgttcggttgcgtttagcaacgcgcgcagcaggaacacgcgctg	2033
Db	1924	TGCTTGGCAGCGCGCTGGGGCTTGGCGAGAGACTGTGGACTGAGAATGGATGCACGT	1983
Qy	2034	gccttgcacgcgaactggccttcttcaaccacgaatcaattgcaactgaactgga	2093
Db	1984	GCCAGTGTCCAGATGCGGCTCTGCTTACCATAGCTTGCACCTGCACAGCTGCATGCCATGGGC	2043
Qy	2094	aatggaagccgcgtgacatgaacgcattatgatggcactattcgaagcaacatggct	2153
Db	2044	AAATACACAGAGGGCACCTCATTTGCTTAACCTTGCACCTGAGTGCCTCGAACCTGGCC	2103
Qy	2154	gaagtggcgcacaattactatgcacgcgcgcgcagaccattctgcatacgcgtaatgacgcg	2213
Db	2104	GAGTGGCGAGAGATGCTATCCATGGCAACGCTGGCAGAGATCTATGTGGCTGCTGCC	2163
Qy	2214	ttcgaataaacgcgcagtcgcccaaaatgttgcaacagttcttccgcgatactacatg	2273
Db	2164	CTGAGGCTCAAGACCAGTCTCCCAAGAGCTTTG---CACTTTTTCACAGTTCTTCTCCTG	2220
Qy	2274	agcgggctcgtcaagagtgcggtgcactaggccacgcgcagcaaacgcgcagagctacgt	2333
Db	2221	AGTAGTCCCGCCAGGCCCTGCCCTGGCACAGAGTGCTCAGTGC---CTCTTGCCATCGAC	2277
Qy	2334	tgggaattcaacgcctatggatatgcctactgcgcgcgcgcgttcttcacgtacgcgt	2393
Db	2278	TGGCTCTGCCACCTCTAGCGCCACCGTTTCTTCTGTGGATGGGGACTGGGGCTGCATG--	2335
Qy	2394	agcgactccggcgcagcagatgattcttcacagctcttaggaatcgaatgataccgcgt	2453
Db	2336	-----GTGCCCCACAGAGAGCCTGTACACGTGGCTGGGAACCCAGTGGATCCCCCTC	2388
Qy	2454	qccacgctcattaaagcaatatgcagagcatttgcgtttaaataccattcagtgctgta	2513
Db	2389	GCCCAGGTGACTCGACTATTCTTGGCAACATCTTTGGAGAGAGACATGAATGTATTGCT	2448
Qy	2514	ggagcggggccacaaatcgggagggctgcaccatctctgtcagcgcgcgcgcgcgcgcgc	2573
Db	2449	CAACCCAGCCCGGGGACACTGATGGAGACAGGGAGTTCTCTGACGCACCTTGGATACCTG	2508
Qy	2574	ttcacgaacagcagcacgcgcgcacattgtcacgaatgttcttaagtaacagtcctc	2633
Db	2509	CAGTTGCTAAATCGCTGCTCATGCTGTGCGGGACTCTGCTCGCAGCTTCTCTGTGTCAGC	2568
Qy	2634	cttaaggacaactctctggctgatgagatgagcgggatacaaacgtgggtgtgtggggcc	2693
Db	2569	TCCAG-----CATGGCTTCCACCACCGGCACAGACCAGTGGCCAGTGTGTGGCC	2619
Qy	2694	gatgttttggagaccgcaglgcacttggctcttggtagacacgcgtggccgcgcaaatg	2753
Db	2620	TCACTACGGCTGTGTGTATCCACTGGCTGCGCGGGATGAAGAGGACGCTGACGGCCTA	2679
Qy	2754	tacgcgcagatcaagcaaatgccacgcgcgtgcacaagtgcgcgcgcgcgcgcgcgcgcgc	2813
Db	2680	TACCCCTGGTAGAGCGTATGCCCAACGCTGTTCAGAGAGACTGAGACACCC-----CTG	2733
Qy	2814	cccaaggcgtgcgtgctgtgcgcgactaagatgatacttactaaaaaacaatggcaac	2873
Db	2734	CCCAAGGCACCTCTGTACTTCCTTCAAGGCTGCCCGGGCTCTGCTGGACCACAGAAAGTG	2793
Qy	2874	gcactggacaaaagtctcaagcaattgtgtaaacatctctctgcgatgagtcgagtgtgag	2933
Db	2794	GAGTGTGCCCAAGCCCTGGCCATCTGTGAGAAGGCCACGCGGTACTTCCGGGACAGC	2853
Qy	2934	ctcaagagtctgttgcgtgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	2993
Db	2854	TTAGCCGCTCCACCAACTGGCAGCTCCATTTGCAAGGCCATGCAGCTGCTCTCTGTGTGAT	2913

QY	2994	cagttgcttaccctgcgattggcgtctcgaataacaggactgcctctgtggaaactggaacac	3053
DG	2914	CTACTTCTTGTCGCCCGCACTAGTAGTGTGGCAGCCACAGCAGTACCAGCCCTCAGCCAG	2973
QY	3054	atgaatatggaagacgatggcttctaccaagtgccagtggaagtgaagtctcgagaagttccag	3113
DG	2974	GTAGCTCACAGTGCCACGAATGGATTCTCAGGCCTCCGCTTTGGAGCTTCGAGGTTTCCA	3033
QY	3114	accgattggaactcgttgcgcaacatgtggagaataataccgaacgccccaatcgcgcata	3173
DG	3034	CAGGACCTTGAGCAGCCTGAGGCGCTTGCGCACAGAATCTCCGGCCTGCTATGAGGAGAGTG	3093
QY	3174	tatttgtacgagcagttgttcgcctgatggctgagcctcacccgtgccaacgcaaac	3233
DG	3094	TTCCTACACGAGGCCACAGCTCGGTGTATGCGAGGGCAAGTCTGCCCGGACACACCAG	3153
QY	3234	ctcttggacagagatctgcgatacgcgaacgcgccacctgcctcatcttctcggcgcagcag	3293
DG	3154	CTCCTGGACCGAAGTCTGCGGAGGGCGGCCGCTCCAGTGGCAAAGGAGGCACTGTAGCT	3213
QY	3294	gaLcgcgcgcgcagcagaactcgtggcggagagcgggaacgggcttcggccatglacgtg	3353
DG	3214	GAGCTGGAGCCTC-----GAGCCACATGCGCGGAGCACACAGAGCCCTTGCTGCTG	3264
QY	3354	gectgcaagtattcccgcctgcgctgcctcagctcccgggtgaaagtgtgtagcatgtta	3413
DG	3265	GCCTCTGCTATCTGCCACCTGCCCTTCCTGTCTGCGCCCCCTGGACAGCAATGAGCATGTTG	3324
QY	3414	gcgcgggcgcacaagcactgagagagtgggcgacagcgaagaaactcaagagagtcttac	3473
DG	3325	GCTGAGGCAGCAGGCACATGTAGAGAAGCTTGTGTATCATCGGCTACTGCTTACTGCTGCCAG	3384
QY	3474	cagctgatgaagtgcctgggcaacggca	3501
DG	3385	CAGATGCTTCTCGCGCTGGCGCGTGGGA	3412
RESULT 15			
RATADDLA			
LOCUS	RATADDLA	3115 bp	mRNA
DEFINITION	Rat addl mRNA sequence.		
ACCESSION	L16995		
VERSION	L16995.1	GI:294509	
KEYWORDS	adipocyte protein; cell determination; cell differentiation; helix-loop-helix protein. Rattus sp. cDNA to mRNA. Rattus sp.		
SOURCE			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
REFERENCE	1 (bases 1 to 3115)		
AUTHORS	Kim, J., Tontoz, P., and Spiegelman, B.		
TITLE	Addl: a novel helix-loop-helix protein associated with adipocyte determination and differentiation		
JOURNAL	Mol. Cell. Biol. 13, 4753-4759 (1993)		
MEDLINE	93330269		
FEATURES	Location/Qualifiers		
source	1..3115 /organism="Rattus sp." /db_xref="taxon:10118"		
BASE COUNT	629 a	1002 c	870 g 614 t
ORIGIN			

Query Match 2.8%; Score 105.8; DB 12; Length 3115;
Best Local Similarity 45.7%; Pred. NO. 1.9e-14;
Matches 586; Conservative 0; Mismatches 662; Indels 33; Gaps 5;

Accession	Sequence	Length
Qy 1029	aagcgtcgcccaacagccatcgagcgcgctatcgacctcaatcaacgacaaagt	1088
Db 925	AAGCGCAGCGCCACATGCCATTGAGAGCGCTACCTTCCTTATCAATGACAAGATT	984
Qy 1089	aacdaattcaagaaactttgtatggaagagcgagcgccaaagctgaacaaatccgcagtta	1148

Db 985 GTGGAGCTCAAGGACCTGTGGTGGGCACTGAGGCAAGCTGAATAAATCTGCTGTCTTG 1044
Qy 1149 cggaaatacagacagagattcgggattctcgaacgcagagatcacgattcgaaggcagag 1208
Db 1045 CGAAGGCCATCGACTACATACCCCTCTTTACAGCACAGCAACAGAACTCAAGCAGGAG 1104
Qy 1209 ttgcagcgccctgcagagggagctaaaggcagcgagcgctcccaagggtgaaggatttactt 1268
Db 1105 AACCTGACCTT-----GCCAAGTGTCTCAGAAAGCAAAATCACTGAAAGA-----C 1149
Qy 1269 cagctgggcaactcgctgtagagcatccaagaagcgccgcgagagctgcagaccttt 1328
Db 1150 CTGGTGTACGCTTGTGGCAGCTGAGGAGGCACAGATGTGTCTATGGAGGCGCATGAAACCT 1209
Qy 1329 accacggatgcggagcagcgccgcagcgagcgatgaatcggatccttcgctcgcgc 1388
Db 1210 GAAGTGGTAGAAACGCTGACCCCTCCACCTCAGACGCGGCTCACCCCTCCAGAGTAGC 1269
Qy 1389 atgcactcggacatctcgtttgcgcgcacatccctatggtggtatccaccgcgcagctgtgc 1448
Db 1270 CCCTTTGCTTGGGCAGCAGCAGCAGCAGTGGTGGCAGTGACTCTGAGCCCGCAGCAGC 1329
Qy 1449 agtggcagcagcagcagcaatgaagaccactggtggtggtccagctctatgcgcggcag 1508
Db 1330 CCAGCCTTTGAGGATAACCAAGGTGAAAGCCAGCGGCTGCTTACATAGCCGAGGCATG 1389
Qy 1509 gccaccaactctcgctcgagctctcatgtttatgttcgcacatcctggccgtcaatccc 1568
Db 1390 CTGGACCCGCTCCCGCTGGCCCTGTGTACTGTCTTCTCTGTCTGACTGCAACCCCA 1449
Qy 1569 ttcaagacctttccagcgcgccactatgacagtaataagcagatcttggcgacatgagc 1628
Db 1450 TTGGCCTCACTGTTGGCTGGGCGATCCTCACTCCCTCTCTGATGCTCGGGTGTGCACCGT 1509
Qy 1629 ggtcaagacgactctctcttacgacgtggaaggtgaagg-----tttgtgtgtgg 1682
Db 1510 AGTTCTGGGGCAGCATGCTGGAGGCGGAGCAGCAGATGGCTTAATTGGACCCAGTGG 1569
Qy 1683 cagcagagttccctggatataagctatgaacttcacactgatgttggtgagcttgggtgaaa 1742
Db 1570 TTGCTGCCACCCCTAGCTCTGGCTGGGCAATGGACTACTAGTGTGGCCTGCTTGGCTCTT 1629
Qy 1743 ttgctggtttacggtgatcc-----gcagctggaagcgcaaacggagcgctactgc 1793
Db 1630 CTCTTTGCTACGGGGNACTGTACACGAGCCACACTCCGGCCGGCTGTACACTTCTGG 1689
Qy 1794 cagcaagcagcggcgctgactctctatttagccaaggacagctgtctcaggcctacgcc 1853
Db 1690 AGACATCGCAAAACAAGCTGACCTGGATTGGCCCGGGAGATTTTGCCACAGCGCGCTCAA 1749
Qy 1854 ggttacctcaactgtctgcatagttttgatttaagtctaccgcgctgcgcttggaggtg 1913
Db 1750 CAGCTGTGGCTGGCCTTGCAGGCCCTGGGCGGCGCCCTGCCACCTCAAACTGGATCTG 1809
Qy 1914 tacttgcaaccacgtggcagttcctcgtttctttccatcgctcctggctgggtcg 1973
Db 1810 GCCTGCAGCCTGCTTTGGAACCTCGTCCGCCACCTGCTGCGAGCGTCTTTGGGTGGGCGC 1869
Qy 1974 gtgctgtcacggcggtccggtgggtgttttagcaacgcccgccagcaggaacacaggcgctg 2033
Db 1870 TGGCTGGCAGGCCAGGCTGGGGGCTCTCAGAGGGGACTTACAGGGTGTGAGAAAGGATGCTCGT 1929
Qy 2034 qcatctgcagcgaactggccctgctcttcaaccgactgaatcaattgcaactgactgga 2093
Db 1930 GCCAGTCCCCGAGATGGCGCTGCTGTACCATTAAGCTGCACCAGCTGCATGCCATGGGC 1989
Qy 2094 aatggaagcgcggtgacatgaacggcatatgatggcactatttcgaagcaacatggct 2153
Db 1990 AAGTACACAGAGGCCATCTTGTGCTTCTTAACCTGGCACTGAGTGGCTTAACCTGGCT 2049
Qy 2154 gaagtggcgcaaatctactgacacgcgcgagaccaatctgcatccacgtaatgacagcg 2213

Db 2050 GAGTGTCCAGGAGATGCTATATCCATGGCAACACTGGCAGAGATCTACGTGGCAGCTGCC 2109
Qy 2214 ttgcgaatgaagcgagtgccccaataatggttgaacagttcttcgcccgcatactacatg 2273
Db 2110 CTAAGGGTCAAAACCCAGCCTCCCCAGAGCCTTGCAC----TTCCTGACACGTTTCTTCTTA 2166
Qy 2274 agccgggctcgtccaagagtgc 2294
Db 2167 AGTAGTGGCCGCCAGGCGCTGC 2187

Search completed: September 4, 2000, 01:11:57
Job time: 23737 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:58:07 ; Search time 299.13 Seconds
(without alignments)
3151.552 Million cell updates/sec

Title: US-09-332-522B-7

Perfect score: 3768

Sequence: 1 gttattaaagctgcaaatat.....aattatttaacagattttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	165	4.4	4154	1	Q79037	Human SREBP-1a cod
2	114.4	3.0	5197	1	Q79038	Human SREBP-2 codi
3	48.4	1.3	234	1	Q84832	Spinocerebellar at
C 4	47.2	1.3	2744	1	Q98470	MisPl-containing p
5	47	1.2	203	1	V17226	SCA2 gene CAG repe
6	47	1.2	203	1	V30271	Glutamine rich reg
7	45.8	1.2	195	1	Q84831	Spinocerebellar at
C 8	44.8	1.2	32207	1	V73805	KSHV LUR DNA (nucl
C 9	44.8	1.2	137507	1	V19941	KSHV long unique c
10	42.2	1.1	543	1	Q23092	Antigen tc-7a gene
11	41.8	1.1	165	1	V30274	Glutamine rich reg
12	41.2	1.1	14704	1	Q20685	PKS 741 insert con
13	41	1.1	171	1	Q84834	Spinocerebellar at
14	41	1.1	1498	1	X23749	Rice anthranilate
15	40.6	1.1	154	1	Q84835	Spinocerebellar at
16	40.6	1.1	168	1	Q84833	Spinocerebellar at
C 17	40	1.1	110000	1	X20248_00	Borrelia burgdorfe
C 18	39.8	1.1	1542	1	X16151	Human Sox1 encodin
19	39.8	1.1	3879	1	V21060	cDNA encoding the
20	39.8	1.1	3879	1	V23009	Human polyhomeotic
21	39.8	1.1	10266	1	T33007	Mouse SRY-related
22	39.2	1.0	726	1	X39871	Gastric cancer ass
C 23	38.8	1.0	786	1	T46148	T. cruzi LL9E homo
C 24	38.2	1.0	50341	1	V22674	DNA sequence of a
C 25	38.2	1.0	52297	1	T51411	Mycobacteriophage
C 26	38.2	1.0	52298	1	Q47357	L5 mycobacterioph
27	38	1.0	506	1	Q84804	Spinocerebellar at
28	38	1.0	3366	1	Q84803	Spinocerebellar at
29	38	1.0	4126	1	X13277	Enterococcus faeca
30	38	1.0	10660	1	Q84793	Spinocerebellar at
31	37.4	1.0	2462	1	V35473	Human hSK3 coding
32	37.2	1.0	1538	1	Q21732	DNA encoding a RAP
C 33	37.2	1.0	3211	1	V05541	Actinobacillus pie

34	37.2	1.0	3284	1	V31337	Mortierella alpina
35	36.8	1.0	245	1	T75504	P. americanus skin
36	36.8	1.0	4356	1	Q37543	Cardiac adenyllyl c
37	36.8	1.0	4356	1	Q95540	Cardiac adenyllyl c
c 38	36.8	1.0	114955	1	X53491	Human adenosine A1
39	36.6	1.0	228	1	Q23094	Antigen tc-10a gen
40	36.6	1.0	1239	1	Q61445	Lignin peroxidase
41	36.6	1.0	1666	1	Q61443	Lignin peroxidase
42	36.6	1.0	1810	1	O51010	Lignin peroxidase
43	36.6	1.0	1810	1	Q61444	Lignin peroxidase
44	36.6	1.0	2051	1	X28069	Heparin sulphate 6
45	36.6	1.0	2287	1	Q92617	Human afamin cDNA.

ALIGNMENTS

RESULT	1
Q79037	
ID	Q79037 standard; DNA: 4154 BP.
AC	Q79037; 1995 (first entry)
DT	04-AUG-1995
DE	Human SREBP-1a coding sequence.
KW	Probe: sterol regulatory element binding protein; regulatory protein; cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW	SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW	transcription factor; low density lipoprotein; LDL; receptor;
KW	3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW	hypercholesterolaemia; enhancer; ss.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
cds	167..3610
FT	/tag= a
FT	/product= SREBP-1a
PN	W09426922-A.
PD	24-NOV-1994.
PF	13-MAY-1994; U05300.
PR	13-MAY-1993; US-061697.
PR	01-OCT-1993; US-131365.
PA	(TEXA) UNIV TEXAS SYSTEM.
PI	Briggs MR, Brown MS, Goldstein JL, Wang X;
DR	WPI; 95-006813/01.
DR	P-PSDB; R66390.
PT	New sterol regulator element binding protein - used to develop
PT	prods. and screening assays for agents for reducing plasma
PT	cholesterol levels (Eng)
PS	Claim 31; Page 185-98; 305pp; English.
CC	The sequences given in Q79037-38 encode the sterol regulatory element
CC	binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC	proteins which are involved in the regulation of genes involved in
CC	cholesterol metabolism that are under the control of an associated
CC	sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC	fail into two families, SREBP-1 and SREBP-2. Both proteins are members
CC	of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC	transcription factors. Each have the ability to bind to SRE sequences
CC	and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
CC	found in the promoters for the low density lipoprotein (LDL) receptor
CC	and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
CC	transcription in the absence of sterols and is inactivated when sterols
CC	accumulate. Human SREBP-2 contains 1141 amino acids and has 47% identity
CC	with SREBP-1a. SREBP-1a was the first recognised member of this family
CC	and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2
CC	includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
CC	identical), and an unusually long extension of 740 amino acids on the
CC	C-terminal side of the bHLH-Zip region. SREBP-2 possesses one feature
CC	lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC	SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor.
CC	production in the presence of sterols. SREBP identified in screening
CC	assays, may be used to reduce plasma cholesterol levels and in
CC	controlling hypercholesterolaemia and its associated diseases.
SQ	Sequence 4154 BP; 714 A; 1406 C; 1280 G; 754 T;

Query Match 4.4%; Score 165; DB 1; Length 4154; Best Local Similarity 45.7%; Pred. No. 1.3e-35; Matches 1144; Conservative 0; Mismatches 1295; Indels 62; Gaps 14;									
QY	1023	gaagtaaacgctcgcccaacacgcacatcgagcgcgctatcgcaacctcaatacaacgac	1082						
DB	1130	GGAGAAAGCGCAGCGCCACACGGCATTTGAGAAGCGCTACCGCTCTCCATCAATGAC	1189						
QY	1083	aagattaacgagttgaagaactggtagtgagagcagcgaagctgaacaagtccgca	1142						
DB	1190	AAATCATTTAGCTCAAGATCTGGTGGTGGGCACGTGAGGCAAGCTGAATAAATATGCT	1249						
QY	1143	gtgttcggaataccatagacaagattcgggattctgcaacgcgcagaaatcacgattcgaag	1202						
DB	1250	GTCTGGCAAGCGCATCGACTACATTCGCTTTCTGCAACACACCAACCAACTCAAG	1309						
QY	1203	gcagagttgcagcgctcgagagggagcttaattggcagcgacgcgctcccaaggtgaaggat	1262						
DB	1310	CAGGAGAACCTAAGTGTGCGCACTGTCTCCACAAAAGCAAAATCTCTGAAGGATCTGGTG	1369						
QY	1263	t----tacttcagctgggactcgctggtagagatccaaagaagccgcgagagctc	1318						
DB	1370	TCGGCTGTGGCATGGAGGGAACACAGACGTGCTCATGAGGGCGTGAAGACTGAGGTG	1429						
QY	1319	gcagacctttaccacgagatgccggactgacgcgcgcacgcagcgatgaatcggatcctc	1378						
DB	1430	GAGGACACATGACCCCAACCCCTCGGATGCTGGTTCACCTTTCAGAGCAGCCCCCTTG	1489						
QY	1379	gctctgcccacgactcgacatctcgttgcgcgcacatccctatggttgatcccaaccgc	1438						
DB	1490	TCCTTTGGCAGCAGGGGCACTGGCAGCGGTGGCAGTGGCAGTGACTCGGAGCGCTGACAGC	1549						
QY	1439	cagctgtagcagtgccagcagcagcagaatgaagaaccaactggtggtgcccagctctat	1498						
DB	1550	CCAGTCTTTTGGAG-GACAGCAAGGCAAGCCAGCAGCAGCGGCCGCTCTGACACAGC----	1603						
QY	1499	gcgcgcgcatggccacccactctcgctcggaactcgatctgcatgtttatgttgcacatcctggc	1558						
DB	1604	-CGGGGATCGTGGACGCTCCCGCTGGCCCTGTGCACGCTCGTCTCTCTGCTGCTGTC	1662						
QY	1559	cgtcaatcccttcaagacctttctcagcgcggccactatgcagtaatgacgacttgg	1618						
DB	1663	CTGCAACCCCTTGCTCTCTTTGTGGGGGCCCGGGGCTTCCAGAGCCCTTCAGATACCAC	1722						
QY	1619	cgcactgagc-----ggtcaagacgcatctctcttctacgagtggaagtgaaag	1669						
DB	1723	CAGGCTACCATAGCCCTGGGCGCACAGCTGTGGGCACCGAGAGCAGATGGCCCTGG	1782						
QY	1670	ttttgtgttcgcagcagagttcctggatggctatggctattgaaacttcacactgatgttg	1729						
DB	1783	CTGGGCCCACTGGCTGTCGCCCCAGTGTCTGCTGCTCAATGGGCTGTTGGTGTCTGT	1842						
QY	1730	atgcttggtaaatgctggtttacggtgatac-----gcagctggacgcgcaaac	1780						
DB	1843	CTCCTTTGGTCTCTCTCTTTGTACGCTGAGCGCAGTGCACACGCGCCCACTCAGGCCCCGC	1902						
QY	1781	ggacgcctactgcagcacagcagcggtcgacttctatttagcaagacagtcgctc	1840						
DB	1903	CGTGTACTCTGAGGCGATCGCAAGCAGGCTGACCTGGACCTGGCCCGGGGAGACTTTGC	1962						
QY	1841	tcagggctacgcgggtttacctcaactgtctgcatagtttggattaaagtacgcgcgc	1900						
DB	1963	CCAGGCTGCCACAGCTGTGGTGGCCCTGCGGGCACTGGCGCGGCCCTTGCCCACTC	2022						
QY	1901	gcgcttgaagtgttacttgcagaccacgctggcagttctctcgtttttttccatgcct	1960						
DB	2023	CCACCTGGACCTGGCTTTGTAGCTCTCTCTGGAACCTCATCCGTCACCTGTGACAGCTCT	2082						
QY	1961	ctggctgggtcggtgtctgcagcggttcggtgggtgttttagcaacgccccagcag	2020						
DB	2083	CTGGGTGGCCCGCTGCTGCGACGCGCGGAGGGGCGCTGCAGGAGGACTGTGCTGCG	2142						
QY	2021	gaacaggcgctggcatctgcacgcgaactggccctgctcttccaaccgactgaatcaatt	2080						
DB	2143	AGTGGATGCTAGCGCCAGCGCCCGAGACGCGAGCCCTGGTCTACCAATAAGCTGCACCAGCT	2202						
QY	2081	gcaactgactggaataatggaagccgcggtgacatgaacgagcattatgatggcactattcgc	2140						
DB	2203	GCACACATGGGGGAACACACAGCGCGGACCTCTACTGCCACCAACCTGCGCGTGAATGC	2262						
QY	2141	aagcaacatggtggaagtggcgccacaatctactgacacccgcgcgagaccatctgcacca	2200						
DB	2263	CCTGAACCTGGCAGAGTGTGACGGGATGCCGTGTGTGTGGCAGCGTGGCCGAGATCTA	2322						
QY	2201	cgtaatgcagcgcttgcgaatgaagcgagctgccccaaaaatggttcaacagttcttcgc	2260						
DB	2323	TGTGGCGGCTGCATTGAGAGTGAAGACCACTCTCCACGGGCGCTTG---CATTTTCTGAC	2379						
QY	2261	ccgatactacatgagccggtcgtcaagagtgcgtcgactagggccacccgagcaaac	2320						
DB	2380	ACCTTTCTTCTCAGGAGTGGCCCGCAGGCTCTGCTGGCACAGAGTGGCTCAGTGC---C	2436						
QY	2321	gcaggagctacgtttgggcattcacagcctatgatatacgctactgcccacgcacgactctt	2380						
DB	2437	TCCTGCCATGCACTGGCTGTGCCACCCCGTGGCCACCCGTTTCTTCG---TGGATGGGGA	2493						
QY	2381	caogtacgattcagcgactcccgcgagcaggatgattcttcacacgctcttaggaatcc	2440						
DB	2494	CTGGTCCGTGCTCAGTACCCCATG-----GGAGAGCCTGTACAGCTTGGCCGGGAACCC	2547						
QY	2441	atgtgatcccgctgcccaogctattaaagcaatatcagagacatttgcgttttaaatccat	2500						
DB	2548	AGTGGACCCCTCGCCACAGGTGACTCAGCTATTTCGGGGAACATCTCTTAGAGGAGCACT	2607						
QY	2501	tcagttcgtgtaggagcggggcacaaaatcgagagcgctgcacacatcttctctcagcgg	2560						
DB	2608	GAAGTGTGTACCCAGCCCAACCCACCCCTGGGTACGTGATGGGGACAGGAATTTCTC	2667						
QY	2561	agaggcggaacagtttcagcaacagcagcagcgccacccattgttcagcaatgttcttaa	2620						
DB	2668	GGATGCCCTCGGTACTGTGAGCTGTGTAACAGCTGTCTCTGATGCTCGCGGGGCTCTGCG	2727						
QY	2621	gtacacgtccctccttaagagacactctctgggtgatgaggtgatgagcgggatacaaacgt	2680						
DB	2728	CTACAGCTTCTCCATCAGTTTCCA---GCATGGCCACACACCCGGGTGATAGCCGGTGGC	2784						
QY	2681	ggtgtggtggcgcatgttttgagacgcgagtcgactggtcctcttgtgtaagacacgct	2740						
DB	2785	CAAGTGTGGGCTCTCTGACAGCTGTGGTGATCCACTGTGCTGGCGGGGATGAGGAGGC	2844						
QY	2741	ggccgagcaatttacgcgaggatcaagcaaatgcccaacgcagctgcacacagtcgcgcga	2800						
DB	2845	GGCTGAGCGGCTGTGCCCGTGGTGGAGCACCTGCCCGGGGTGCTGCAGGAGCTTGAGAG	2904						
QY	2801	aaacgatcatctcccaagcgctgcatgctgtgctgcgagctaaagatgatcttactaaa	2860						
DB	2905	ACCCCTGCCCAGGCG---AGCTCTGCACCTCTTCAAGGCTGCCCGGGCCCTGCTGGGCTG	2961						
QY	2861	aaacaatggcaacgcagctggacaaaagtctcaagcaattgttaaacatcctctgcgatga	2920						
DB	2962	TGCCAA-GGCAGAGTGTGTTCCAGCCAGCCTGACCATCTGTGAGAGGCCAGTGGGTACC	3020						
QY	2921	gtcagtggtgagctcccaagagtgttgcactgtaaacccggatcacacgcgcacccagggtat	2980						
DB	3021	TGCAGGACAGCCTGGCTTACCACACCAG--CCAGCAGCTCCATTTGACAAGCCGCTGCAGCT	3078						
QY	2981	aaagctgcttttcagttgcttacctgcgcatgtgctgctcgaataagtagactgctgtg	3040						
DB	3079	GTTCCTGTGTGACTGCTCTTCTGTGTGGCCACAGCAGCTGTGGCGGACGAGCAGCCCCC	3138						
QY	3041	ggaaatggaacacatgaatatgagagacgatggcttcttaccagtgccagtgtaagtgcct	3100						
DB	3139	GGCCCGGCCCCAGCAGCCCGCAGGGCACACAGCAGCGGCCCCAGGCTTCCGCCCTTGAGCT	3198						
QY	3101	cgagaagttccagacgcgatttgaaactcgttgcgcaacattgttggagaataataccgaacgc	3160						

Db 3199 GCGTGGCTTTCCACAGGGAGCTGAGCAGCCTGAGCGGCTGGCACAGAGCTTCGGGCCCGC 3258
Qy 3161 ccaatcgccatatatttagagagcagtttgcctctgatggctggagcctcaccgtg 3220
Db 3259 CATCGGAGGGTGTCTCTACATGAGGCCACAGGCCCGGCTGATGGCGGGGCGAGCCCCAC 3318
Qy 3221 tccaacgcaacagctcttgagcagagtgctgcgacacgcccactcgtccatctt 3280
Db 3319 ACGGACACACAGCTCTCTGACCGAGTGTGAGCGCGCGGCGAGCCCGGTTGGCAAGG 3378
Qy 3281 ctgcggaagaaggatcgggcgagcagaactctgtggcgagagcggaagcggtctc 3340
Db 3379 AGCGCGGTTGGCGAGCTGGAGCCGC-----GGCCACGCGCGGCGAGCACGCGGA 3429
Qy 3341 gqccatgtacgtggcctgcaatgtctccgcctgcgtctgcagctccccgggtgaacg 3400
Db 3430 GGCCTTGTCTGGCTCTCTGCTACCTGCCCCCGGCTTCTCTGTGCGGCGCGGCGAGCG 3489
Qy 3401 tgctggcatgttagcggagcgccgaagaccctggagaaggtggcgacaagcgaaagct 3460
Db 3490 CGTGGCATGCTGGCTGAGCGCGCGGCACACTCGAGAAGCTTGGCGATCGCGGCTGCT 3549
Qy 3461 caagagtgctaccagctgatgaagtcgctgggcaacggca 3501
Db 3550 GCACGACTGTACGACAGATGCTCATGCGCTGGCGCGGTGGGA 3590

RESULT 2
Q79038
ID Q79038 standard; DNA; 5197 BP.
AC Q79038;
DT 04-AUG-1995 (first entry)
DE Human SREBP-2 coding sequence.
KW Probe: sterol regulatory element binding protein; regulatory protein;
KW cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW transcription factor; low density lipoprotein; LDL; receptor;
KW 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW hypercholesterolemia; enhancer; ss.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 118..3543
FT cds
FT /*tag= a
FT /product= SREBP-2

WO9426922-A.
PN 24-NOV-1994.
PP 13-MAY-1994; US-061697.
PR 01-OCT-1993; US-131365.
PA (TEXA) UNIV TEXAS SYSTEM.
PI Briggs MR, Brown MS, Goldstein JL, Wang X;
DR WPI: 95-006813/01.
DR P-PSDB: R66391.
PT New sterol regulator element binding protein - used to develop
PT prods. and screening assays for agents for reducing plasma
PT cholesterol levels (Eng)
PS Claim 34: Page 219-26; 305pp; English.
CC The sequences given in Q79037-38 encode the sterol regulatory element
CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC proteins which are involved in the regulation of genes involved in
CC cholesterol metabolism that are under the control of an associated
CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC transcription factors. Each have the ability to bind to SRE sequences
CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
CC found in the promoters for the low density lipoprotein (LDL) receptor
CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
CC transcription in the absence of sterols and is inactivated when sterols
CC accumulate. Human SREBP-2 contains 1141 amino acids and has 47% identity
CC with SREBP-1a. SREBP-1a was the first recognised member of this family
CC and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2
CC includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%

CC identical), and an unusually long extension of 740 amino acids on the
CC C-terminal side of the bHLH-zip region. SREBP-2 possesses one feature
CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC SREBP promotes SRE-1-mediated gene transcription. eg. LDL receptor
CC production in the presence of sterols. SREBP identified in screening
CC assays, may be used to reduce plasma cholesterol levels and in
CC controlling hypercholesterolemia and its associated diseases.
SQ Sequence 5197 BP; 1066 A; 1570 C; 1434 G; 1127 T;
Query Match 3.0%; Score 114.4; DB 1; Length 5197;
Best Local Similarity 52.3%; Pred. No. 1.7e-21;
Matches 315; Conservative 0; Mismatches 266; Indels 21; Gaps 2;
Qy 2909 cctctgcgtagtgagtgagagctccaagagtgcttgactgtcaacccggatcaccca 2968
Db 2946 CCATTCCGAGAGGGCAGTGGCCACCCTATGGAGCAGCCTCAAGCTCAGTGGGGGACCTC 3005
Qy 2969 cgcgaagggtataaaagctgcttttccagttgcttaccgtgcgattggctgcgaactag 3028
Db 3006 TGACCTTGGCTTCAACACAGCTGGTCCAGCTGCTCACTGTGACCTGCTACTGTGCTACG 3065
Qy 3029 gactgctctgtgggaactggaaacacatgaatatggaggacgatgg---cttctaccaagt 3085
Db 3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGGCAGCCAGGCTGTGGGGGAGACCTTACCACGC 3125
Qy 3086 gccagtggaagtgcgagaagttccagaccgatttgaactcgttgcgaacatttgga 3145
Db 3126 GTCAGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCGAGCTGCCAGGCTGGCACCA 3185
Qy 3146 gaatataccgaacgcccgaatcgcgcatatatattgtacgagcgagtttgcgcctgatggc 3205
Db 3186 CAGCTTCCGCCCCAGCATACCCGAAGGTGTTCTTCGATGAAGCCACCCTGCGCTGATGGC 3245
Qy 3206 tggagcctcacogtctccaacgcaacagctctttggacagagtgcttgatcgcgaacgc 3265
Db 3246 AGGAGGACGCCCCACCCGACCCACCCAGCTGCTGGAACACACAGCTGCGGGCGGCACAC 3305
Qy 3266 ccactgcgtccattcttgcggcagaaggatcgcgcgacagaactctcgtggcgagaga 3325
Db 3306 GCA-----GAGCACCAAGCAGGAGGTGGATGCTTGGCCCGGCCCA 3347
Qy 3326 gcgggaaocggggttcggccatgtacgtggcctgcaagtatctccgcctgcgctgcag 3385
Db 3348 GCGAGAGCGGGCCACCGCCATCTCTGTGCGCTGCCGCCACCTGCCCTCTCTCTCTCTC 3407
Qy 3386 ctccccgggtgaactgctggaatgttagccgagggcgccaaagaccctggaagtggg 3445
Db 3408 CTCCCGGGGCCAGCGGGCAGTGTGCTGGCGGAAGCTGCCCGCACCTTGGAGAAGGTGGG 3467
Qy 3446 cgacaagcgaagctcaaggagtgctaccagctgatgaagtcgctgggaacggcattgg 3505
Db 3468 CGACCGCGCGCTCTCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGTGGCTACGTC 3527
Qy 3506 ca 3507
Db 3528 CA 3529

RESULT 3
Q84832
ID Q84832 standard; DNA; 234 BP.
AC Q84832;
DT 25-SEP-1995 (first entry)
DE Spinocerebellar ataxia type 1 CAG repeat region patient #2.
KW Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW CAG repeat region; patient #2; ss.
OS Homo sapiens.
PN WO9501437-A.
PD 12-JAN-1995.
PF 29-JUN-1994; U07336.
PR 29-JUN-1993; US-084365.
PR 28-JUN-1994; US-267803.


```
RESULT 6
V30271
ID V30271 standard; DNA; 203 BP.
AC V30271;
DT 02-OCT-1998 (first entry)
DE Glutamine rich region encoding sequence found in SCA2 patients.
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease; ss.
OS Homo sapiens.
PN WO9818920-A1.
PD 07-MAY-1998.
PF 30-OCT-1997; J03946.
PR 30-OCT-1996; JP-304059.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S.
DR WPI; 98-272215/24.
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
PS Example 1; Page 22; 38pp; Japanese.
CC This represents a sequence encoding glutamine rich repeat region which
CC can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC specification provides a gene sequence causative of the neurodegenerative
CC disease SCA2, having a tri-nucleotide (CAG) repeat region which in the
CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
SQ Sequence 203 BP; 68 A; 70 C; 62 G; 3 T;
```

```
Query Match 1.2%; Score 47; DB 1; Length 203;
Best Local Similarity 52.9%; Pred. No. 0.0014;
Matches 101; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 316 acctggcgcgcagcagatgtacaacatgctgtgacagcagctgaacgcataccacgc 375
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5 ACCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 64

QY 376 aacacgagtcctgtgagcagcccaatccctcagcaaacacgcgcagctgaaagcg 435
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 65 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 124

QY 436 agcactcttcgccagtcacatcaaggaggaactgcatcagcagcaacaacagtcgcgc 495
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 184

QY 496 ttctcgtctac 506
|| | | | |
DB 185 ATCACGGAAAC 195

RESULT 7
Q84831
ID Q84831 standard; DNA; 195 BP.
AC Q84831;
DT 25-SEP-1995 (first entry)
DE Spinocerebellar ataxia type 1 CAG repeat region patient #1.
KW Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW CAG repeat region; patient #1; ss.
OS Homo sapiens.
PN WO9501437-A.
PD 12-JAN-1995.
PF 29-JUN-1994; U07336.
PR 29-JUN-1993; US-084365.
PR 28-JUN-1994; US-267803.
PA (MINU ) UNIV MINNESOTA.
PI Chung M, Orr HT, Zoghbi HY;
```

```
DR WPI; 95-061001/08.
PT New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT - used to develop prods. for detection or presymptomatic
PT diagnosis of a SCA1 disorder
PS Disclosure; Fig 2; 11pp; English.
CC Q84831-Q84835 show the CAG repeat regions of five individuals
CC affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC CAG repeat region (Q84804) that the mutations responsible for
CC SCA 1 occur. The full nucleic acid (Q84793) and its protein product
CC (K71111) can be used to develop products, for the presymptomatic
CC detection of a SCA 1 disorder.
SQ Sequence 195 BP; 60 A; 67 C; 64 G; 4 T;

Query Match 1.2%; Score 45.8; DB 1; Length 195;
Best Local Similarity 52.3%; Pred. No. 0.003;
Matches 101; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 329 gcagatgtacaacatgctgtgacgagcctcgaacacataccacgacgacgcagtcgcgt 388
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 62

QY 389 ggatcagcagccgcgaatccgtcgcagcaacagccgcgcagctgaaaagcagcactctcgcc 448
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 122

QY 449 agtgcacatcaaggaggaactgcatcagcagcaaacacagtcgcgcgttctcgtctacaa 508
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 182

QY 509 accagatccccctc 521
|| | | | | | |
DB 183 TCCGGGGCTCATC 195

RESULT 8
V73805/C
ID V73805 standard; DNA; 32207 BP.
AC V73805;
DT 25-FEB-1999 (first entry)
DE KSHV LUR DNA (nucleotides 105,301-137,507).
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;
KW v-adi; G-protein coupled receptor; FGARAT; ds.
OS Kaposi's sarcoma-associated herpesvirus.
PN US5849564-A.
PD 15-DEC-1998.
PF 29-NOV-1996; 770379.
PR 29-NOV-1996; US-770379.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR WPI; 99-069741/06.
PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
PS Disclosure; Column 155-182; 109pp; English.
CC This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for ORF65 which encodes capsid protein IV, ORF67
CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC ORF59, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,
CC ORF73 which encodes immediate early protein (IEP), K14 which encodes
CC OX-2 (v-adi), ORF74 which encodes G-protein coupled receptor, ORF75
CC which encodes tegument protein/FGARAT, K15, KSHV is a new human
CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC most common form of neoplasm occurring in persons with acquired immune
CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC and for detecting expression of a DNA virus associated with Kaposi's
CC sarcoma in a cell.
SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T;
```

```

Query Match      1.2%; Score 44.8; DB 1; Length 32207;
Best Local Similarity 53.4%; Pred. No. 0.12; 82; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 321 gcgcgcagcagatgtacaacatgctgtgagcagcctcgaaacgcataccagcagcagc 380
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19807

QY 381 cagtcgcgtggtatcagcagcgcgcgaatccgtcgagcaacagccgcacgtgaaagcgagcac 440
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 19806 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAG 19747

QY 441 tcttcgcagtgacatcatcaggaggaaactgcatacagcagcaacaaacagtcgcgcct 496
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 19746 GATGAACAGGAGCAGCAGGAGGAGCAGCAGCAGCAGGAGGAGCAGGAGGAGGATTT 19691

RESULT 9
V19941/c
ID V19941 standard; DNA; 137507 BP.
AC V19941;
DE 03-AUG-1998 (first entry)
KW KSHV long unique coding region and terminal repeat.
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW complement-binding protein; glycoprotein; capsid protein IV; infection;
KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW HIV immune status; anti-inflammatory agent; therapy; ds.
KW Kaposi's sarcoma-associated herpes virus.
FH Key Location/Qualifiers
FT CDS 1142..2794
    /tag= a
    /product= complement-binding protein
    /tag= b
    /product= glycoprotein B
    /tag= c
    /product= interleukin 6
    /tag= d
    /product= macrophage inflammatory protein II
    /tag= e
    /product= interferon regulatory factor 1
    /tag= f
    /product= protein T1.1
    /tag= g
    /product= glycoprotein M
    /tag= h
    /product= glycoprotein L
    /tag= i
    /product= interferon regulatory factor 2
    /tag= j
    /product= interferon regulatory factor 3
    /tag= k
    /product= glycoprotein X
    /tag= l
    /product= interferon regulatory factor 4
    /tag= m
    /product= capsid protein IV
    /tag= n
    /product= immediate early protein

```

```

FT FT
PN PN
PD PD
PR 22-JUL-1997; U13346.
PR 29-NOV-1996; US-757669.
PR 25-JUL-1996; US-686243.
PR 25-JUL-1996; US-686349.
PR 25-JUL-1996; US-686350.
PR 25-JUL-1996; US-687253.
PR 25-JUL-1996; US-688814.
PR 05-SEP-1996; US-708678.
PR 10-OCT-1996; US-728323.
PR 13-NOV-1996; US-747887.
PR 13-NOV-1996; US-748640.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR WPI; 98-130615/12.
PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
PS Example 2; Page 135-203; 230pp; English.
CC This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral INF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC while the protein can be used in protective vaccines. Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, leukaemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC can be inhibited with methotrexate. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP III may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.
CC Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T;
SQ

```

```

Query Match      1.2%; Score 44.8; DB 1; Length 137507;
Best Local Similarity 53.4%; Pred. No. 0.29;
Matches 94; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 321 gcgcgcagcagatgtacaacatgctgtgagcagcctcgaaacgcataccagcagcagc 380
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 125166 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 125107

QY 381 cagtcgcgtggtatcagcagcgcgcgaatccgtcgagcaacagccgcacgtgaaagcgagcac 440
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 125106 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAG 125047

QY 441 tcttcgcagtgacatcatcaggaggaaactgcatacagcagcaacaaacagtcgcgcct 496
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 125046 GATGAACAGGAGCAGCAGGAGGAGCAGCAGCAGCAGGAGGAGGAGGAGGATTT 124991

RESULT 10
Q23092
ID Q23092 standard; DNA; 543 BP.
AC Q23092;
DT 17-AUG-1992 (first entry)
DE Antigen tc-7a gene.
KW Oocysts; MAB 12-07; sporozoite; ss.

```

OS *Elmeria tenella*.
PN WO9204460-A.
PD 19-MAR-1992.
PR 05-SEP-1991; U06430.
PR 12-SEP-1990; US-581693.
PA (GENE-) GENEX CORP.
PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI Ruff MD, Augustine PC, Danforth HD;
DR WPI; 92-114365/14.
DR P-PSDB; R22392.
PT Vaccine against avian coccidiosis - comprising recombinant
PT *Elmeria* antigen ac-1b or ac-6b gene, or microorganisms expressing
PT them
PS Claim 5; Page 39 + Fig 5; 56pp; English.
CC To identify antigens of *E. tenella*, expression libraries were
CC prep'd in lambda vector, lambda gt11, using cDNA prep'd. from polyA
CC mRNA isolated from *E. tenella* oocysts. The cDNA expression
CC library was screened with monoclonal antibody (Mab) 12-07 which was
CC raised against the sporozoite stage of *E. tenella*. The library
CC to be screened was plated on a host that allows lysis and plaque
CC formation. During induction of the antigens encoded by the phage,
CC the plaques were identified by screening the filters with Mab 12-07.
CC The cDNA inserts from the Mab 12-07 positive phage were cloned into
CC bacteriophage M13 and subjected to sequence analysis. Following
CC sequence analysis, *E. tenella* antigen tc-7a was identified.
SQ Sequence 543 BP; 229 A; 135 C; 161 G; 18 T;

Query Match 1.1%; Score 42.2; DB 1; Length 543;
Best Local Similarity 49.8%; Pred. No. 0.055;
Matches 107; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 274 acctgcactcggacatgctggacatcctcaacacatggagacctggcgccgacacaga 333
Db 153 AAGCAGCAGCTGCAACAGCAGCAGCTCAGCAACACCACTGCGAGCTGCAGCACCACTGC 212

QY 334 tgtacaacatgctgctggagcagcctcgaacgcataccacgacgacgagcctccgtggatc 393
Db 213 AGCAGCACCAGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 272

QY 394 agcagccgaacatccgtcgagcaacacgcccgcactgtaaaagcgagcactcttccagtcg 453
Db 273 GACAGCCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 332

QY 454 acatcaagagggaactgcacacgacgacgacgacgacgacgacgacgacgacgacgacgac 488
Db 333 AGCAGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 367

RESULT 11
V30274
ID V30274 standard; DNA; 165 BP.
AC V30274;
DT 02-OCT-1998 (first entry)
DE Glutamine rich region encoding sequence found in SCA2 patients.
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease; ss.
OS Homo sapiens.
PN WO9818920-A1.
PD 07-MAY-1998.
PF 30-OCT-1997; J03946.
PR 30-OCT-1996; JP-304059.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DR WPI; 98-27215/24.
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
PS Examples: Page 23; 38pp; Japanese.
CC This represents a sequence encoding glutamine rich repeat region which
CC can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC specification provides a gene sequence causative of the neurodegenerative
CC disease SCA2, having a tri-nucleotide (CAG) repeat region which in the

CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
SQ Sequence 165 BP; 55 A; 55 C; 55 G; 0 U;

Query Match 1.1%; Score 41.8; DB 1; Length 165;
Best Local Similarity 53.3%; Pred. No. 0.035;
Matches 88; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 330 cagatgtacacatgctgctggagcagcctcgaacgcataccacgacgacgacgacgacgacgac 389
Db 1 CAG 60

QY 390 gatcagcagccgaatccgtcgagcaacacgcccgcactgtaaaagcgagcactcttcacca 449
Db 61 CAG 120

QY 450 gtgcacatcaagagggaactgcacgacgacgacgacgacgacgacgacgacgacgacgacgac 494
Db 121 CAG 165

RESULT 12
Q20685
ID Q20685 standard; DNA; 14704 BP.
AC Q20685;
DT 22-APR-1992 (first entry)
DE PKS 741 insert contg. Sry.
KW Y-chromosome; SRY; sex-determining region; genital ridge; ss.
OS Mus musculus.
PN WO9200375-A.
PD 09-JAN-1992.
PF 28-JUN-1991; G01057.
PR 28-JUN-1990; GB-014446.
PR 13-JUL-1990; GB-015488.
PR 09-MAY-1991; GB-010085.
PA (IMCR) IMP CANCER RES TECH.
PI Goodfellow PN, Lovell-Badge R;
DR WPI; 92-041557/05.
PT Nucleic acid pY53.3, its fragments and oligo-nucleotide(s) - is
PT used for determining or controlling sex of e.g. embryos by
PT detecting Y chromosomes
PS Claim 7; Page 122 and Fig 19(1-3); 183pp; English.
CC The mouse gene equivalent to the human Y-located gene defined by
CC pY53.3 (Q20684), is referred to as Sex-determining Region of the
CC Y-chromosome (Sry) (formerly (mating type box Y) MtY).
CC The mouse Sry is expressed in male genital ridge, at the time that
CC sex determination occurs, and adult testis.
CC See also Q20680-81, Q20684-87, Q22530-31, Q22535-36 and Q22560-67.
SQ Sequence 14704 BP; 4259 A; 3237 C; 3099 G; 4103 T;

Query Match 1.1%; Score 41.2; DB 1; Length 14704;
Best Local Similarity 49.1%; Pred. No. 0.76;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 274 acctgcactcggacatgctggacatcctcaacacatggagacctggcgccgacacacaga 333
Db 8802 ACCACCACCCAGCAGCAGCAGCAGCTTCTATGACCACCACCCAGCAGCAGCAGCAGCAGCAG 8861

QY 334 tgtacaacatgctgctggagcagcctcgaacgcataccacgacgacgacgacgacgacgac 393
Db 8862 AGCAGCAGCAGCAGCTCCATGACCAACACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8921

QY 394 agcagccgaacatccgtcgagcaacacgcccgcactgtaaaagcgagcactcttcaccaatgc 453
Db 8922 AGCAGCAACAGCAGCTCCATGACCAACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8981

[illegible]

Search completed: September 4, 2000, 01:05:08
Job time: 18483 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:33:20 : Search time 183.89 Seconds
(without alignments)
2818.589 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gttattaaagctgcaaatat.....aatatttaacagattttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/pctus_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	165	4.4	4154	1	US-08-131-365B-37
2	165	4.4	4154	3	US-08-668-123-37
3	114.4	3.0	5197	1	US-08-131-365B-53
4	114.4	3.0	5197	3	US-08-668-123-53
5	48.4	1.3	234	2	US-08-469-802B-3
6	48.4	1.3	234	3	US-08-267-803B-3
c 7	47.2	1.3	2793	2	US-08-209-747-1
c 8	47.2	1.3	2793	2	US-08-458-298-1
9	45.8	1.2	195	2	US-08-469-802B-2
10	45.8	1.2	195	3	US-08-267-803B-2
11	44.8	1.2	3489	4	US-08-728-323A-1
c 12	44.8	1.2	32207	3	US-08-770-379-20
c 13	44	1.1	7218	1	US-08-232-463-14
14	42.2	1.1	543	7	5273901-6
15	41.6	1.1	7218	1	US-08-232-463-14
16	41	1.1	171	2	US-08-469-802B-5
17	41	1.1	171	3	US-08-267-803B-5
18	40.6	1.1	154	2	US-08-267-803B-6
19	40.6	1.1	154	3	US-08-469-802B-6
20	40.6	1.1	168	3	US-08-469-802B-4
21	40.6	1.1	168	3	US-08-267-803B-4
22	40	1.1	533	7	5482709-5
c 23	38.8	1.0	786	2	US-08-403-379A-2
c 24	38.8	1.0	786	4	US-08-929-414-2
25	38.8	1.0	793	4	US-09-018-595B-1
26	38.8	1.0	1187	2	US-08-440-856A-2

c 27	38.2	1.0	3252	4	US-08-809-740A-1	Sequence 1, Appli
c 28	38.2	1.0	3252	4	US-08-809-740A-4	Sequence 4, Appli
c 29	38.2	1.0	50341	2	US-08-247-901C-1	Sequence 1, Appli
c 30	38.2	1.0	50341	4	US-09-075-904-1	Sequence 1, Appli
31	38	1.0	3366	2	US-08-469-802B-1	Sequence 1, Appli
32	38	1.0	3366	3	US-08-267-803B-1	Sequence 1, Appli
33	38	1.0	10660	3	US-08-267-803B-8	Sequence 8, Appli
34	37.2	1.0	506	2	US-08-469-802B-7	Sequence 7, Appli
35	37.2	1.0	506	3	US-08-267-803B-7	Sequence 7, Appli
36	37.2	1.0	1519	1	US-07-971-759-19	Sequence 19, Appl
37	36.6	1.0	228	7	5273901-10	Patent No. 5273901
38	36.6	1.0	228	7	5482709-9	Patent No. 5482709
39	36.6	1.0	2287	1	US-08-222-619-1	Sequence 1, Appli
40	36.6	1.0	2287	6	PCT-US95-04075-1	Sequence 1, Appli
41	36.4	1.0	1995	2	US-08-425-069-3	Sequence 3, Appli
42	36.4	1.0	1995	4	US-08-317-844B-3	Sequence 3, Appli
43	36.2	1.0	802	4	US-09-018-595B-2	Sequence 2, Appli
c 44	36.2	1.0	4615	1	US-08-188-582-1	Sequence 1, Appli
c 45	36.2	1.0	4615	1	US-08-646-715-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-131-365B-37
; Sequence 37, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 167...3607
US-08-131-365B-37

Query Match 4.4%; Score 165; DB 1; Length 4154;
Best Local Similarity 45.7%; Pred. No. 8.7e-38;
Matches 1144; Conservative 0; Mismatches 1295; Indels 62; Gaps 14;

QY	1023	gaagtaaagcgtcggcccaaacagccatcagcggcgctatcgcaacctcaatacaacgac	1082
DB	1130	GGAGAAAGCGACAGCCCCACAAAGCCATTGAGAAGCGCTACCGCGCTCCTCCATCAATGAC	1189
QY	1083	aagattaacgagctgaagaacttgtagtgagagcagcgccaagctgaacaagtcgca	1142
DB	1190	AAATCATTTGACTCAAGATCTGCTGTGGGCACCTGAGGCAAAAGCTGAATAAATCTGCT	1249
QY	1143	gtgttcggaaatccatagacaagattcgggatctgcaacgccagaatacagatctgaag	1202
DB	1250	GTCTTTGGCAAGGCCATCGACTACATTCCGCTTTCTGCAACACAGCAACCAAGAACTCAAG	1309
QY	1203	gcagagtgcagcgcctcgaagagagctaatggcacgcgcgctccaaggtgaagat	1262
DB	1310	CAGGAAAGCTTAAGTCTGCGCATGCTGTCCAAAAAGCAAAATCTTGAGGATCTGGTG	1369
QY	1263	t----tacttcagctgggcaactcggccctggttagagcatccaagaagcgcgcgagagctc	1318
DB	1370	TCGGCGTGTGGCAGTGGAGGAACAACAGACGCTGCTCATGTGAGGGCGCTGAAGACTGAGGTG	1429
QY	1319	gcagaccttaccacgagtcggaagctacgcgccacgcagcagatgaatcgatcttc	1378
DB	1430	GAGGACACTGACCCCAACCCCTCGGATGCTGGCTCACCTTTCAGACGACGCCCGCTTG	1489
QY	1379	gtctctgcacctgcactcggacatctgttcgcgcacatccactatggtggatccaacgc	1438
DB	1490	TCCTTTGGCAGCAGGGGAGTGGCAGCGGTGGCAGTGCAGTACTCGGAGCCTTGACAGC	1549
QY	1439	cagctgtagcagtgcacgcagcagcagcaatgaagaacacactggtggtgccagctctat	1498
DB	1550	CCAGTCTTTGAG-GACAGCAAGGCAAAAGCCAGAGCAGGGCGCGTCTCTGCACAGC	1603
QY	1499	ggcggcattggccacccactctcgcctcggactctgcatgttatgttcgccactcctggc	1558
DB	1604	-GGGGCATGCTGGACCGCTCCCGCTGGCGCTGTGCAGGCTCGCTCTTCCTTCGCTGCTC	1662
QY	1559	cgtcaatcccttcaagacctttctccagcgcgcgcactatgacagtaatgacgatcttg	1618
DB	1663	CTGCAACCCCTTGGCCTCTCTGCTGGGGGCCCGGGGCTTCCAGCCCCCTCAGATACCAC	1722
QY	1619	cgacatgagc-----ggtcaaaagcgaattctcttcaacgctggaaggtgaag	1669
DB	1723	CAGCGCTTACCATAGCCCTGGGCGCAAGCTGTGGGACCCAGAGCAGAGATGGCCCTGG	1782
QY	1670	ttttgctgtctggacagagttcctgatatgctattgaacttcaactgatcttg	1729
DB	1783	CTGGGCCCCAGTGGCTGCTCCCCCAGTGGTCTGGCTGCTCAATGGGCTGTTGGTGCTGT	1842
QY	1730	atgcttggtgaatctggttttaacggtgatcc-----gcagctggacgcgcaaac	1780
DB	1843	CTCCTTTGGTCTCTCTTTGTCTACGGTGAGCCAGTCAACGGGCCCACTCAGGCCCCGC	1902
QY	1781	ggacgctactccacagcagcagcagcggtcgaactctattttagcaagacagctcgtc	1840
DB	1903	CGTGTACTTCTTGAGGCATPCGCAACAGCGTGAAGCTGGACCTGGCCCGGGAGACTTTGC	1962
QY	1841	tcaggcctacgcggttacctcaactgctcgaatggtttggttaagtataccgcgctc	1900
DB	1963	CCAGGCTGCCACGACGCTGTGGCTGTGGCGCTCGGGGCATGCGGCGCGGCCCTGCCACCTC	2022
QY	1901	gccttggagtgtaacttgcaaacacagtggaagttccttcgttttcttttcacatgcct	1960
DB	2023	CCACCTTGGACTCGGCTGTGTAGCCTCTCTTGAACCTCATCCGTCACCTGCTGCTGACGCTC	2082
QY	1961	ctggctgggtcgggtgctgtcaacgcggtcccggtgggctgttttagcaacgcgcgcagcag	2020
DB	2083	CTGGGTGGCCGCTGGCTGGCAGCGCGGGCAGGGGGCGCTGCAGCAGGACTGTGTCTGTGCG	2142

Qy		2021	gaaacaggcgctggcactctgcagccggaacctggccctgtcttccaacagcaactgaatcaatt	2086
Db	2143	AGTGGATGCTAGCGCCGAGCGCCGAGACGACGCCCTGTGTTACCATAAAGCTGCACCAAGCT		2202
Qy		2081	gcaactgactgaaaatggaagccgcggtgacatgaacgcgcatatgatggcacatttcgc	2140
Db	2203	GCACACCATGGGAAGCACACAGGCGGGCACCCTACTGCCACCACCTGGCGCTGAGTGCC	2262	
Qy		2141	aagcaacatggctgaagtggccacaactactacacccgcgcgcgcagaccactctgcccca	2200
Db	2263	CCTGAACCTGGCAGAGTGTSCAGGGGATGCCGTGTGTGGCGACGCTGGCGAGATCTA	2322	
Qy		2201	cgtaatgacagcgttgccgaatatgaagccgcagtgccccaaaaatggtgcacaagttcttcgc	2260
Db	2323	TGTGGCGGCTGTCATTGAGAGTAGAACACCAGTCTCCACCAGGGCTTGG---CAATTTCAGAC	2379	
Qy		2261	ccgatacatagccgcggctctgcaaagtgccgactagggcgcgcagcagcaaac	2320
Db	2380	ACGCTTCTTCTGAGCAGTGGCCGCCAGGCCCTGCCTGGCACAGAGTGGCTCAGTGC---C	2436	
Qy		2321	gcaggagctacgtttggcattcaacagcctatggatatcgctactgcgcgcgcgcacgtcttt	2380
Db	2437	TCCTTGCCATCAGTGGCTGTGCCACCCCCTGGCGCACCGTTCTTCG---TGGATGGGGGA	2493	
Qy		2381	cacgtacgactgtagcgactccgcgcgcgcagcagatggattcttcacagctottagaataccc	2440
Db	2494	CTGCTCGCTGCTCAGTAACCCCATG-----GGAGAGCCCTGTACAGCTTGGCGCGGAAACCC	2547	
Qy		2441	atgtgatccccgctgcccgaogtcattaaagcaatcagcagagcaattgcgttgtttaatccat	2500
Db	2548	AGTGGACCCCTGGCCGACGCTGACTCAGCTATTTCGGGAACAATCTCTTTAGACGACGACCT	2607	
Qy		2501	tcaagtctcgttagcagcgggcgcacaaatcgcgagggcctgcgccacatctctgcacgcgg	2560
Db	2608	GAACTGTGTGACCCAGCCCAACCCCCAGCCCTGGGTACGCTGATGGGGACAGGAATTCTC	2667	
Qy		2561	agagcgggaacagttgcagcaacagcagcacagcgcgcacccattgtcagcaatgttcttaa	2620
Db	2668	GGATGCCCTCCGGTACTGTCAGCTGCTGAACAGCTGTTCTGTATGTGCGGGGGCTCCTGC	2727	
Qy		2621	gtacaagtcctccttaagagcaactctctggcctgatgaggatgagcgggatacaaacgt	2680
Db	2728	CTACAGCTTCTCCATCAGTTTCCA---GCATGGCCACACACCACCGCGTACACCCCGTGGC	2784	
Qy		2681	ggctgggggcgcgatgtttggagaccgcagtcgactggctccttggctgaagacaagct	2740
Db	2785	CAAGTGTGGGCCTCTGACAGCTGTGGTGATCCACTGGCTGCGCGCGGATGAGGAGGC	2844	
Qy		2741	ggccgagcaattbtacgcgcaggatcaagcaaatgcccaocgcagctgcacaacagtcgcgcga	2800
Db	2845	GGCTTGAGCGGCTGTGCCCGCTGTGTGAGGACACCTGCCCGGGTGTCTGACAGAGTCTGAGAG	2904	
Qy		2801	aaacgatacatctcccaagcgcctcatcgtctgctgcagctaaagatgatcttactaaa	2860
Db	2905	ACCCCTGCCAGGGC---AGCTCTGCATCTCTTCGAAGGCTGCCCGGGCCCTGCTGGGGCTG	2961	
Qy		2861	aaacaatggcaacgcactggcaaaaaagctcagcaaatggtaaacatctctctcgcgatga	2920
Db	2962	TGCCAA--GGCAGAGTCTGTGTCTCCAGCCAGCCTGACCATTGTGAGAAGGCCAGTGGGTACC	3020	
Qy		2921	gtcgagtgtagcctcaagagtgcttgactgtcaacgcgatccgcgcgcgcgcgcgcgcgcgcgc	2980
Db	3021	TGCAGGACAGCCTGGCTTACCACACCAAG--CCAGCAGCTTCCATTGACAAGGCCCGTGCAGCT	3078	
Qy		2981	aaagctccttttcagttgcttacctgcgattggctgtctcgaaactaggaactgcctctgtg	3040
Db	3079	GTTCCCTGTGACTGCTTCTTGTGTGTGGCGCACACGACCTGTGGCGGCACGACGACGCCCC	3138	
Qy		3041	ggaactggaacacaatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgcct	3100
Db	3139	GGCCCCGGCCCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	3198	
Qy		3101	cgaagaattccaadccqatttgaactcgtttgcgcacaactttggaggaataaccggaacgc	3160

;
; LENGTH: 5197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-131-365B-53

Query Match 3.0%; Score 114.4; DB 1; Length 5197;
Best Local Similarity 52.3%; Pred. No. 5.8e-23;
Matches 315; Conservative 0; Mismatches 266; Indels 21; Gaps 2;

QY 2909 cctctgcgatgagtgagtggtgagctccaagagtgcttgagcttcaacccggatcaccca 2968
DB 2946 CCATTGCCGAGAGGCGCAGTGGCCACCTATGGAGCAGCCCTCAACGTCAGTGGGGGCACCTC 3005

QY 2969 gcccaagggtataaagctgtctttccagttgcttacctgcgattgctgcgaactag 3028
DB 3005 TGACCCCTGCCCTCAACCAACGCTGTCAGCTGCTCACTGTGACCTGCTACTGCGCTACG 3065

QY 3029 gactctctgtgggaacttgaacacatgaatatgaggagcagtg---cttctaccaagt 3085
DB 3066 GACACGCGCTCTGGCAAAACAGGCCAGTGGCCAGCGCTGTGGGGGAGACCTACCAACGC 3125

QY 3086 gccaggtgaagtgtcgcagaagttccagaccgatttgaactcgtttgcgcgaacattgtga 3145
DB 3126 GTACGCGCGCTGAACCTGGCGGGCTTCCACGGGACCTGGCGACCCCTGCGGAGCTGGCACA 3185

QY 3146 gaataaccgaagcccaatcgcgcataattttgtacgagcgagtttgcgcctgatggc 3205
DB 3186 CAGCTTCCGCCCCAGCATACCGCAAGGTGTCTCGCATGAAGCCACCGTGGCGCTGATGCG 3245

QY 3206 tggagcctcacccgttccaacgcacacagctcttggacagggagctcggatccgcaaacgc 3265
DB 3246 AGGAGGCGAGCCCCACCCCGACCCAGCTGTGGAACACAGCCCTGCGGCGCGCACAC 3305

QY 3266 ccactcgtccattcttctgcggcagcaagatcggggcgagcaagaacttcgtggggcgaga 3325
DB 3306 GCA-----GAGCACCAAGCAGGAGGTGATGCTGCGCCCGGCCCA 3347

QY 3326 gggggaacgggttcggccatgtacgtggcctgcgaagtatctcccgctgcgctgcag 3385
DB 3348 GCGAGAGCGGGCCACGGCCATCTCTGCGCCTCCCGCCACCTGCGCCTTCCTTCCTCTC 3407

QY 3386 ctccccgggtgaactgctggtgcatgttagccgagcgcgccaagaccctggagaagtgag 3445
DB 3446 GCGAGCGCGCTCTCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

RESULT 4
US-08-668-123-53
; Sequence 53, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee

;
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-668-123-53

Query Match 3.0%; Score 114.4; DB 3; Length 5197;
Best Local Similarity 52.3%; Pred. No. 5.8e-23;
Matches 315; Conservative 0; Mismatches 266; Indels 21; Gaps 2;

QY 2909 cctctgcgatgagtgagtggtgagctccaagagtgcttgaactgtcaacccggatcaccca 2968
DB 2946 CCATTGCCGAGAGGCGCAGTGGCCACCTATGGAGCAGCCCTCAACGTCAGTGGGGGCACCTC 3005

QY 2969 gcccaagggtataaagctgtctttccagttgcttacctgcgattgctgcgaactag 3028
DB 3006 TGACCCCTGCCCTCAACCAACGCTGTCAGCTGCTCACTGTGACCTGCTACTGTCGTACG 3065

QY 3029 gactcgtctgtgggaacttgaacacatgaatatgaggagcagtg---cttctaccaagt 3085
DB 3066 GACACGCGCTCTGGCAAAACAGGCCAGTGGCCAGCGCTGTGGGGGAGACCTTACCAACGC 3125

QY 3086 gccaggtgaagtgtcgcagaagttccagaccgatttgaactcgttgcgaacattgtga 3145
DB 3126 GTACGCGCGCTGAACCTGGCGGGCTTCCACGGGACCTGGGCGAGCTGGCGAGGCTGGCACA 3185

QY 3146 gaataaccgaagcccaatcgcgcataattttgtacgagcgagtttgcgcctgatggc 3205
DB 3186 CAGCTTCCGCCCCAGCATACCGCAAGGTGTCTCTGATGAAGCCACCGTGGCGCTGATGCG 3245

QY 3206 tggagcctcacccgttccaacgcacacagctcttggacagggagctcgcgatccagcaacgc 3265
DB 3246 AGGAGGCGAGCCCCACCCCGACCCAGCTGTGGAACACAGCCCTGCGGCGCGCACAC 3305

QY 3266 ccactcgtccattcttctgcggcagcaagatcggggcgagcaagaacttcgtggggcgaga 3325
DB 3306 GCA-----GAGCACCAAGCAGGAGGTGATGCTGCGCCCGGCCCA 3347

QY 3326 gggggaacgggttcggccatgtacgtggcctgcgaagtatctcccgctgcgctgcag 3385
DB 3348 GCGAGAGCGGGCCACGGCCATCTCTGCGCCTCCCGCCACCTGCGCCTTCCTTCCTCTC 3407

QY 3386 ctccccgggtgaactgctggtgcatgttagccgagcgcgccaagaccctggagaagtgag 3445
DB 3446 GCGAGCGCGCTCTCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

RESULT

; ZIP

Query Ma

RESULT 7

; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
US-08-232-463-14

```
Query Match      1.1%; Score 41.6; DB 1; Length 7218;
Best Local Similarity 3.1%; Pred. No. 0.12;
Matches 11; Conservative 197; Mismatches 146; Indels 0; Gaps 0;

Qy 554 gcgcagcggcctttaaagccgcgcccaacacagccaccatcacatgacgacca 613
      | || | | | || | | | | | | | | | | | | | | | | | | | | | |
Db 1049 GTCGAGGGAGCTGCGATYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 614 gcggatgccgcgaacacagcggtgtatcccccatctctggcagtagcttgtacca 673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1109 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 674 gtccatgtcccccgcacgtcgcgggtggagtctgcgaaccagaatgtcaatgtcagca 733
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1169 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 734 gcccggtgtgcgaactcctctgcccgcttctgtctccttggccccagcagtcgatacgca 793
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1229 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 794 acccttcattacgtacaactctaagcgcggaatgaacttcgatgaagcccatgtacttgtct 853
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1289 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 854 cttagcagccacggtagccagtcacaccccatctccacctgtggctccaccacc 907
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1349 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1402
```

Search completed: September 4, 2000, 00:36:23
Job time: 20353 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:25:16 ; Search time 4457.97 Seconds
(without alignments)
3727.309 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gtttattaagctgcaaatat.....aataatttaacacagattttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
59: gb_est33: *
60: gb_est34: *
61: gb_est35: *
62: gb_est36: *
63: gb_est37: *
64: gb_est38: *
65: em_est27: *
66: em_est28: *
67: em_est29: *
68: em_est30: *
69: gb_est39: *
70: gb_est40: *
71: gb_est41: *
72: gb_est42: *
73: gb_est43: *
74: gb_est44: *
75: em_est31: *
76: em_est32: *
77: em_est33: *
78: em_est34: *
79: gb_est45: *
80: gb_est46: *
81: gb_est47: *
82: em_est35: *
83: em_est36: *
84: em_est37: *
85: gb_est48: *
86: gb_est49: *
87: gb_est50: *
88: gb_est51: *
89: gb_est52: *
90: gb_est53: *
91: gb_est54: *
92: gb_est55: *
93: gb_gss1: *
94: gb_gss2: *
95: gb_gss3: *
96: gb_gss4: *
97: em_gss1: *
98: em_gss2: *
99: em_gss3: *
100: em_gss4: *
101: gb_gss5: *
102: gb_gss6: *
103: gb_gss7: *
104: gb_gss8: *
105: gb_gss9: *
106: em_gss5: *
107: em_gss6: *
108: em_gss7: *
109: em_gss8: *
110: em_gss9: *
111: em_gss10: *
112: em_gss11: *
113: gb_gss10: *
114: gb_gss11: *
115: em_gss12: *
116: gb_gss12: *


```
Db 421 TGGCAGTAGCTTTGCTACCAAGTCCATGTCCTCCCGCCACGCTCGCGGTGGAGTTCGGCA 480
QY 712 accaagaatgaatgatcagagccggttgctgaacctcctgctcccgcttctgctcctt 771
Db 481 ACCGAATGTCGAATGTCATGACGCCGCTTGCTGCAACTCCTGCTCCCGCTTCTGCTCCTT 540
QY 772 Tgccccagcagtcgctatccgcgaaccttcattacgtacaaactcctaagccggaatgactt 831
Db 541 TGCCCCAGCAGTCGTATCCGCAACCTTCATTACGTACAACTCTAAGCGCCGAATGACTT 600
QY 832 cogatgaagccatgtacttctgtctcttcgcagcccaacggtag 870
Db 601 CCGATGAAGCCATGTACTTNGCTTTCGACGCCACCGTAG 639

RESULT 2
LOCUS All107040 631 bp mRNA EST 24-NOV-1998
DEFINITION GH06327.5prime GH Drosophila melanogaster head pot2 Drosophila
melanogaster cDNA clone GH06327 5prime similar to U38238: HLH106
FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA sequence.
ACCESSION All107040
VERSION All107040.1 GI:3475975
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 631)
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 63 row: C column: 3
High quality sequence stop: 539.

FEATURES
Source
1..631
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH06327"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site:1; EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT 200 a 180 c 140 g 111 t
ORIGIN
1..631
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH06327"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site:1; EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."

Query Match 16.7%; Score 631; DB 35; Length 631;
Best Local Similarity 100.0%; Pred. No. 2.1e-160;
Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttataagctgcaataatactcgtgaaaaaaatcaaaacaccatgaacaaagtgtt 61
Db 1 TTTATTAGCTCAAATATACTCGTGAATAAATCAAAACACCACTGAACAAAGTGTT 60
QY 62 gaactattactaactagtcgtagtgttaaaacaaagtcggtgacattaaacagttag 121
Db 61 GCAACTATTACTAAGTAGTCGCTAGTTTAAAGCAAAGTCGCTTGACATTAAACAGTTATG 120
QY 122 gaaaacaaaacacacgcgtgaactaagaaaacagatagaagtggtgaagcattcgcaat 181
|||||
```

/note="Organ: head--brain & sensory organ; Vector: Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 151 a 181 c 157 g 109 t
ORIGIN

Query Match 15.6%; Score 589.4; DB 28; Length 598;
Best Local Similarity 99.8%; Pred. No. 4e-149;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 709 cgaacagaaatgcaatgcatgcagccgctgtctcaactctctccgctctgtctc 768
|||||
Db 1 CGNACCAGAAATGCAATGTCATGCACGCCGTTGCTGCAACTCTGCTCCGCTTCGCTC 60
|||||

QY 769 cttgtcccccagcagtcgtatccgcgaacccttcattcgtacaaactaaaggccggaatga 828
|||||
Db 61 CTTTGGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGCGCGGAATGA 120
|||||

QY 829 cttccgatgaagccatgactgtcttgcagcccaagctagccagtcacaaaccccatctc 888
|||||
Db 121 CTTCCGATGAAGCCATGTACTTCTCTTGACGCCACGGTAGCCAGTCCAAACCCCATCTC 180
|||||

QY 889 cactgtggtctccaccacgcagcaagcacaggtagtcgggcccagcaaggtgcagtgccac 948
|||||
Db 181 CACCTGTGCTCCACCACCGACAGCACAGGTAGTCTGGGCCACCAAGGTGCGAGTGGCAC 240
|||||

QY 949 cactggtccgtcacctgcgcgtctatggaagtcagggccaggaaggtacatcatcaacccgggttc 1008
|||||
Db 241 CACTGCTCCGTCACTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTC 300
|||||

QY 1009 aacccaaggtgaaggaagtaagagctcgtgcccacaaacccatcgagcgctatcgca 1068
|||||
Db 301 AACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCACCAACGCCATCGAGCGCGCTATCGCA 360
|||||

QY 1069 cctcaatcaacgacagattaacgagttgaagaactgtagtgggagagcagggccaagc 1128
|||||
Db 361 CCTCAATCAACGACAGATTAACGAGTTGAAGAACTTGGTAGTGGAGAGCAGGCCCAAGC 420
|||||

QY 1129 tgaacaagtcgcagtggttcggaatccatagacaagattcggggtatctgcaacgcgaga 1188
|||||
Db 421 TGAACAAGTCCGCGAGTGTTCGGGAAATCCATAGACAAGATTCCGGATCTGCAACGCCAGA 480
|||||

QY 1189 atcacatctgaagcagagttcacgcctgcagcgtgcaaggaagctaatggcagcgacggt 1248
|||||
Db 481 ATCACCATCTGAAGGCAGAGTTGCAGCGCTTCGAGAGGGAGCTAATGGCACGCGCGGCT 540
|||||

QY 1249 ccaagtggaaggattacttacttcagctgggcactcgccctggtgtagagcatcca 1299
|||||
Db 541 CCAAGTGAAGGATTTACTTCAGCTGGCAGCTCGGCCTGGTAGAGCATCCA 591
|||||

RESULT 4

AI533148
LOCUS 552 bp mRNA EST 18-MAR-1999
DEFINITION SD04874.5prime SD Drosophila melanogaster Schneider L2 cell culture
POT2 Drosophila melanogaster cDNA clone SD04874.5prime similar to
U38238: HLH106 FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA
sequence.

ACCESSION AI533148.1 GI:4447283

VERSION EST

KEYWORDS

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.

1 (bases 1 to 552)

Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,

Brokstein,P., Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

REFERENCE

AUTHORS

TITLE

JOURNAL COMMENT

Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 48 row: G column: 2
High quality sequence stop: 243.

FEATURES source

1..552
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD04874"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT 128 a 190 c 135 g 99 t
ORIGIN

Query Match 14.2%; Score 534.4; DB 40; Length 552;
Best Local Similarity 98.0%; Pred. No. 3.4e-134;
Matches 541; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 519 ctcataggcacaagctacaattgtcccccagcaacagccagggccttttgaaggccgc 578

Db 1 CTCATAGCCACAAGCTACAAATTGTCCCAAGCAACAGCCAGGGCCTTTTGAAGGGCGCC 60

QY 579 caaccaagcaccacatacatcacatggagccacagcgatcgccgaacacagcgagtg 638

Db 61 CAACCAAGCAGCCACCAATACATACATGGAGCCCGGATGCGCGAAGCGGGGTG 120

QY 639 tatcccccactcttgggcagtagctttgtaccagttccatgtcccccgcacgctgcgcg 698

Db 121 TATCCCCCATCTGTGGCAGTAGCTTTGTCTACCACTCCATGTCCCGCCACGTCGCCG 180

QY 699 gtggagtgcgaaccagaaatgcaatgcaatgcagcccggtgtgcaactcctgtccc 758

Db 181 GTGGAGACTGCGAAACCAAAATGGGAATGTATGCAGCCGTTGTCTGCAACTCCTGTCTCC 240

QY 759 gctctgcctcttgcgccagcagtcgtatccgaaccccttcattacgtacaaactcaag 818

Db 241 GCTTATGCTCTTTGGCCCCAGCAGTCGTATCCCAACCCCTTCATTTAGTACAACTTAAG 300

QY 819 gccggaatgacttcgcgatgaagccatgtactgtcttgagccccacggtagccagtcga 878

Db 301 GCCGGAATGACTTCGGATGAAGCCATGTACTTGTCTTGACGCCACGGTAGGCAGTCCA 360

QY 879 accceatctccactgtggtccaccacgacacagcaggtagtcgggcccagcaaggtg 938

Db 361 ACCCCATCTACACTGTGGCTCCACCACCAAGGACAGGTAGTCGGGGCAGCAAGGTG 420

QY 939 cgaagtggcaccactggtcctccgtcacctgcgctatggaagtcaggggcagggtaccctatc 998

Db 421 CGAGTGGCACCACCTGCTGCTGCTCACCCTCGCGGTATGGAAGACCAGGGCAAGGTACCTATC 480

QY 999 aaccgggttcaaccccccaagtggaaggaaagtaaagcgtcggccccacacgcacatcgagcgg 1058

Db 481 AACCGGGTTCAACCCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCAACAATGCCATCGAGCGG 540

QY 1059 cgctatcgacc 1070

Db 541 CGCTATCGCAAC 552

RESULT 5

AI533491

LOCUS

DEFINITION SD05333.5prime SD Drosophila melanogaster Schneider L2 cell culture

pOT2 Drosophila melanogaster cDNA clone SD05333 5prime similar to U38238: HLH106 Fggn0015234 PID:gl079656 SPtREMBL:Q24146, mRNA sequence.

ACCESSION A1533491
VERSION A1533491.1 GI:4447626
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 521)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 53 row: C column: 9
High quality sequence stop: 510.
Location/Qualifiers
1. .521
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD05333"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site.1: EcoRI; Site.2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."

BASE COUNT 115 a 187 c 118 g 101 t
ORIGIN
source

Query Match 13.8%; Score 519.4; DB 40; Length 521;
Best Local Similarity 99.8%; Pred. No. 4e-130;
Matches 520; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 496 ttctgcttacaacacagatccctctatgacacagctacattgtccccagcaacagc 555
|||||
Db 1 TTCTGCTTACAACACAGATCCCTCATAGCCACACGCTACAAATTGTCCTCCACCAACAGC 60
|||||
Qy 556 cgacgggcttttgaaggcgcgcacacacacagccaccatacatcatcgagcccgac 615
|||||
Db 61 CGACGGGCGCTTTGAAGGCGCGCCCAACCAACAGCCACCATACATCATGAGCGCCAGC 120
|||||
Qy 616 ggaatgcgcgaacacagcggtgtatcccccatctctgggcagtagctttgtctaccagt 675
|||||
Db 121 GGATGCGCGGAACACGCGGGGTATCCGCCATCTCTGGCGAGTAGCTTTGTGTACGAGT 180
|||||
Qy 676 ccatgtccccgcgcacgtgcgcggtgagctctgcgaacacagaaatgtcaatgtcatgcagc 735
|||||
Db 181 CCATGTCCCGCCAGTCGCGGTGGAGTCTGCGAACACAGAAATGTCAATGTATCATGCAGC 240
|||||
Qy 736 ccgttgtctcaactcgtcccgcttctgctccttgcgccagcagtgatcccgcaac 795
|||||
Db 241 CCGTTGCTCAACTCCTGCTCCGCGTCTCTGCTCTTGTCCCGCAGCAGTCTGTATCCGCAAC 300
|||||
Qy 796 ccttcattacgtacaactctaaggccggaatgacttccgatgaagccatgacttgtctct 855
|||||
Db 301 CCTTCAATACGTACAACTTAAGCGCGGAATGACTTCCGATGAAGCCATGTACTTGTCTCT 360
|||||
Qy 856 tcagcccgctagccagtcacaccccatctccacctgtgctccaccacgacaagca 915
|||||
Db 361 TGACGCCACGGTAGCGAGTCCAAACCCCATCTCCACTGTGGCTCCACCACCGCAAGCA 420
|||||
Qy 916 caggtagtcgggcccagcaggtgtagtgaggcaccactggctccgctcacctgcccgtatgg 975
|||||

Db 421 CAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACCACCTGGCTCCGTCACCTGCCGCTATGG 480

Qy 976 aagtcaggggcaaggtacctatcaacgggttcaacccaag 1016
|||||

Db 481 AAGTCCAGGGCAAGGTACCTATCAACGGGTTCAACCCAAG 521

RESULT 6
LOCUS AA201358 487 bp mRNA EST 29-NOV-1998
DEFINITION LD04332.5prime LD Drosophila melanogaster embryo BlueScript
Drosophila melanogaster cDNA clone LD04332 5prime, mRNA sequence.
ACCESSION AA201358
VERSION AA201358.1 GI:1797151
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 487)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Blast expect value = 2.5E-185 on U38238: Drosophila melanogaster transcription factor HLH106 mRNA, complete cds
Plate: 43 row: C column: 8
High quality sequence stop: 485.
Location/Qualifiers
1. .487
/organism="Drosophila melanogaster"
/db_xref="BDGP-EST:BDcln003739"
/db_xref="taxon:7227"
/clone="LD04332"
/clone_lib="LD Drosophila melanogaster embryo BlueScript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: BlueScript SK; Site.1: EcoRI; Site.2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in BlueScript SK(+/-)"

BASE COUNT 161 a 121 c 112 g 93 t
ORIGIN
source

Query Match 12.2%; Score 461; DB 22; Length 487;
Best Local Similarity 99.8%; Pred. No. 2.8e-114;
Matches 472; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 gtttattagctgcacataatactactcgtgaaaaaatacaaaacacccatgaacacaagtgt 60
|||||
Db 16 GTTTATTAAAGCTGCAATAATACTCTGTAAGCAAAATCAAAACACCACTGAACAACTGT 75
|||||
Qy 61 tgcaactattactaaactagtcgctagttaaagcaagtcggtgacattaacagttat 120
|||||
Db 76 TGAACATTATTACTTAAGTTCGCTAGTTTAAAGCAAGTCGTTTGACATTAAACAGTTAT 135
|||||
Qy 121 ggaacacaaaagcacacgctgaactaagaacacagatagaagtggttaaagcattcgcac 180
|||||
Db 136 GGAAACAAACACACACGCTGAAGTAAAGAAACAGATAGAAGCTGGTAAAGCATTTCGCA 195
|||||
Qy 181 tggacacacactgatgaacttaataagacgctccgctggacgagtcattgattgtca 240
|||||
Db 196 TGGACACGACACTGATGAACCTTAATAGACGCTCCGCTGGACGAGTCCATGATTTGTTC 255
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:00:24 ; Search time 11551.9 Seconds
(without alignments)
637.752 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 ggtgcctgactgttttcta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 972840 seqs, 892348106 residues 1945680
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_om:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl1:*

8: gb_pl2:*

9: gb_pr1:*

10: gb_pr2:*

11: gb_pr3:*

12: gb_ro:*

13: gb_sts:*

14: gb_sy:*

15: gb_un:*

16: em_fun:*

17: em_hum1:*

18: em_hum2:*

19: em_in:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_sy:*

29: em_un:*

30: em_vi:*

31: gb_htg1:*

32: gb_htg2:*

33: gb_inl:*

34: gb_in2:*

35: em_ba1:*

36: em_ba2:*

37: em_hum3:*

38: em_hum4:*

39: gb_pr4:*

40: gb_htg3:*

41: gb_htg4:*

42: gb_htg5:*

43: gb_htg6:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_pl3:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3279.6	79.4	67188	45	AC005424	AC005424 *** SEQUE
2	3279.6	79.4	135182	34	AC007121	AC007121 Drosophil
3	3279.6	79.4	313072	58	AE003784	AE003784 Drosophil
4	2390.4	57.9	60625	43	AC020090	AC020090 Drosophil
5	1126.4	27.3	12978	42	AC014229	AC014229 Drosophil
6	1017.4	24.6	109699	42	AC009255	AC009255 Drosophil
7	263.2	6.4	790	13	CNS01P6Y	AL154442 Anopheles
8	193.6	4.7	3976	9	D83782	D83782 Human mRNA
9	187.4	4.5	4119	12	CG067060	U07060 Cricetus
10	153.2	3.7	118474	41	AC008339	AC008339 Drosophil
11	137	3.3	109699	42	AC009255	AC009255 Drosophil
12	62.4	1.5	10844	42	AC017232	AC017232 Drosophil
13	62.4	1.5	28380	34	AE003066	AE003066 Drosophil
14	61.4	1.5	16047	42	AC017996	AC017996 Drosophil
15	61.4	1.5	311786	58	AE003781	AE003781 Drosophil
16	61.2	1.5	95668	67	AC009254	AC009254 Drosophil
17	60.4	1.5	40940	33	CED2013	247808 Caenorhabdi
18	60.4	1.5	191908	31	CEY51B9	298947 Caenorhabdi
19	59.6	1.4	40400	42	AC012915	AC012915 Drosophil
20	59.6	1.4	158322	51	AC011759	AC011759 Drosophil
21	59.6	1.4	274456	34	AE003574	AE003574 Drosophil
22	58.8	1.4	8132	42	AC018078	AC018078 Drosophil
23	58.8	1.4	33240	34	AE003021	AE003021 Drosophil
24	57.8	1.4	33013	34	DMU87285	U87286 Drosophila

Qy 1057 gtatttcgttccccgcttttttgccttatgtgctgaatcacacacagcggagcttg 1116
Db 63807 GTATTTCGTTCCCGCTTTTTCGCTGCCTTATGTAGCGTAATCACACAGCGGAGCTTG 63866
Qy 1117 gccatgctcccttgctgtgtttcttcttggcctgacaatttcgctgcagtcacaaagac 1176
Db 63867 GCCATGTCCCTTGGCTTGTGTCTCTTGGCTGACAATTTGCTGCACTCAAGGAC 63926
Qy 1177 attttcccttaccttgtaactcttggtgattgaaatagcttggatcacaaaagc 1236
Db 63927 ATTTTCCTTACCTTTGTAATCTTTGGGTGATTGAAATAGCTTGGTGATCACAAAGAC 63986
Qy 1237 gtatctctcaatggacgagacattcgacgtgaagatccgcgtggcgacgctcttagcaag 1296
Db 63987 GTACTCTCAATGGACGAGACATTCGACGTGAAGATCCGCTGGCGCAGGCTTTAGCAAG 64046
Qy 1297 gagggttggcatatataccaaagactcttttgacggagataacaaatttggacaattggcttt 1356
Db 64047 GAGGGTTGGCATATATCCAAAGACTCTTTTGGACGGAGATAACAATTTTGACAATTTGGTCTT 64106
Qy 1357 gctactttcgtgccgttcacccagagtttttgatatcttttgccatagtcggtctgtcttcc 1416
Db 64107 GCTACTTTCTGGCCCGTCATCCAGAGTTTTGTATCTTTGGCATAGTCGGCTTGTCTTCC 64166
Qy 1417 gattttatgctacagatgctgctctcttccaacaatactggccatgaacattaaagcggacc 1476
Db 64167 GATTTTATGCTACAGATGCTGCTCTTCTCAACAATACTGGCCATGAACATTAAGCGGACC 64226
Qy 1477 gagtatacggcggagccaaagcaccttctctaagatgttgctgagctgcacccaaagggct 1536
Db 64227 GAGTATACGGCGGAGGCCAAGACGCTTCTTAAGATGTGTGCTGAGCTGCACCCAAAGGGCT 64286
Qy 1537 ggtcgacagatttcccatttttcggggccgccccagacactggcacgcttgttccctggc 1596
Db 64287 GGTGCACAGGATTTCCGATTTTTTGGGGCCGCCCCAGACACTGCCACCGTTTTGTCCCTGGC 64346
Qy 1597 acatttcagcgttctcagtcgcataccaaaactgtgttttgctgatcccgcatctgttagc 1656
Db 64347 ACATTTCAGCGTTCTCAGTCGCATCCAAAACGTGTGTTTGTGTATCCCGCATCTGTTAGC 64406
Qy 1657 gatcgtacaaagcttgggttaatggacactcgtcgcggagagaacaaatacccaaacgata 1716
Db 64407 GATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACGAATAACCCAAACGGATA 64466
Qy 1717 aagattgtaaatttctgggcgcgaactcgttttttcagcgtgcttccatgaatcgtatg 1776
Db 64467 ANGATGTAAATTTCTGGGCCGGAGCTCGCTTTTTTCAGCGTGCCTTCATGATCGGATG 64526
Qy 1777 attgttggaLatgctctatagtttataattcgggatatcctggagcagttgttttagcatg 1836
Db 64527 ATTGTGGATATGCTCTATAGTTTATAATTCGGGATATCTGGAGCAGTTGTTTAGCATG 64586
Qy 1837 cagggcaacggcaaatgacggcaaccttgaacttcaacggcgactacagcgggtcgg 1896
Db 64587 CAGAGCAACGGCAATGACGGCAACCCCTTGAACCTTCAACGGCGACTACAGCGCGGTCCG 64646
Qy 1897 ggaacgttcacagcttttttcagggatggcaagcagcgacggcgacgtgcacaaagtccg 1956
Db 64647 GGACAGTCACAGTTTTTTTCAGGGATGGCAAGCGGACGGCGACGTCGCCACGAGTCCG 64706
Qy 1957 ccaagcgaagcggcttttctacgccaaataaaaagctctctagcgatcgatataaacgaa 2016
Db 64707 CCAAGCGGAACGGCTTTTCTACGCCAATAAAGCTCTCTTAGCGATCGATATAACGAA 64766
Qy 2017 acggccggaggaaatgatgagacttogaatccccagcttcgcacctaaactatttctttca 2076
Db 64767 ACGGCCGAGGAAATGATGAGACTTCGATATCCAGTTCAGCTTCGACCTAAACTATTTCCTTCA 64826
Qy 2077 aacttccactggtccacgattatgaacagatcaacaactctcactaagtgggcactacgtt 2136
Db 64827 AACTTCCACTGGTCCACGATTATGAACAGATACAACATCTCACTAAGTGGCACACTACGTT 64886
Qy 2137 acctcgtcaaccgaccattcgccttagtcatgcccacgtcccggtcgcctagccactctgttg 2196

Db 64887 ACCCTGCTACCGACCATTCGCCTTAGTCATGCATCGCTCCCGAGCTACCCACTCTGTTG 64946
Qy 2197 cggaaatccgcaggagcagctgcaacaaaattttcfaatggaaggccctagcgcgtcgaatcc 2256
Db 64947 CGGAATCCCGAGGACAGCTGCACAAAATTTTCAATGGAAGCGCTAGCCGCTGCACATC 65006
Qy 2257 gatccgttggaactttaacgatgaacagcgtggccgtgagctcccgatgtaatgacagag 2316
Db 65007 GATCCGCTGGACTTTAAGCATGACACAGCTGCCCGTGACTCTCCGATGGTAAATGGCAGAG 65066
Qy 2317 gggttgctctcgtttcccaagagcccccatggaaatatttttcggcatcctcttgtctgc 2376
Db 65067 GGGTTGCTCTTGGTTCCCAAGAGCCCCATGGAATATTTTTCGCCATCTCTTGTGCTGC 65126
Qy 2377 atcagcatcttcgtcttgcttgaacagatggtggttttctaccgctgcatatgtaccag 2436
Db 65127 ATCAGCATCTTCGTCTTGTGTACACGATGGTGGTTTTCTACCGTGCATATGTACCAAG 65186
Qy 2437 aactatgccgagtggtcgtccagttggcacgaatctgaggcaacctgacaagcagactga 2496
Db 65187 AACTATGCCGAGTGGCGCTCCAGTTGGCAGCAATCTGAGGCACCGTTACAAAGCAGACTGAG 65246
Qy 2497 caaatcctcggaggaggttccaaacgcaaatcgcggacacaaacatcgcattgaatgcctg 2556
Db 65247 CAATTCCTGGAGGAGTTCCAACGCAATCGCCGGACACAAACATCGCATTTCAATGCCTG 65306
Qy 2557 gtgtcagcgccgtctacataatacagctgctgccttaaggccaaatccagatgtgggat 2616
Db 65307 GTGTCTGACGGCGCTACATATACAGCTGCTGCCTTAAAGGCCAAATCCGAGTGTGGAT 65366
Qy 2617 gcacgcaatgccgacgactaacacagcatcctccgatccgatcogatalcagatctctcagcag 2676
Db 65367 GCACGCACTGGCAGCACCTAACAGCATCTCCCATCCGATATTTCAGATCTCTCAGCAG 65426
Qy 2677 cggacggatggcgacagcgtgtaacgaaagctggcgtgtcacccggtctcgttgctttgac 2736
Db 65427 CGGACGGATGGCAGACGCTGGTACGAAAGCTGGCGGTGTACCCGCTGTCCGCTGCCTTGAC 65486
Qy 2737 tacttcgataatctaataatcgagtaggtgcccacacggccgcgtagaatttgtgggaatcc 2796
Db 65487 TACTTCGATAATCTAATCCAGTAGGCTTCGCCAACGGCGCGTAGAATTTGTGGGAATCC 65546
Qy 2797 cctcgggatctcaagtgtgcataccaggaagcgcggaagaaacacaggtataaac 2856
Db 65547 CCTCGGGAATTGCTTAAAGTGTGCATACCAGGAAGACGCCGAAGAAACACCGGTATAACC 65606
Qy 2857 cacatccacctgaacggcgatcagatgattgtggcgcgtctttaaaggccgactagatttt 2916
Db 65607 CACATCCACCTGAACGGCGATCGAGTGAATTGTGGCGGTCTTAAATGGCGGCTAGATTTT 65666
Qy 2917 taccgcttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttac 2976
Db 65667 TACCCTTTAGAGACGTACTACAAGGGGAAGCAAAATCGACTGGGGTTTTACCTCGGCTTAC 65726
Qy 2977 aggaagaa----- 2983
Db 65727 AGGAGAGTATGTCTTTTCGTTTAAATACCTTTTAAAAAACTCCTTAAAAAGAAAGACTAT 65786
Qy 2984 -----ctcatgttcgaactgatacactggagcctggagctgggatttaattgttcgacagc 3034
Db 65787 TTCCPTCAGCTCATCTTCGAACCTGGATCCACTGGAAGCTGGGATTAATGTTTGCAGCAGC 65846
Qy 3035 agcgtgttcagcaagaacatccccagaagaccacaaaggagaaatgaataatcaactgg 3094
Db 65847 AGCGTGTGACGAAGAAGCATCCAGAAGACCACCAGGAGAAATGAAAAATCACAATTGG 65906
Qy 3095 aggggtgaagactagcccatcagcagcccaatcacatgcatgacgctcgttaacgacatgg 3154
Db 65907 AGGGTGTAAAGACTAGCCCATCAGCAGCAATCACATGTCAGGTGCTTAAACGACATGG 65966
Qy 3155 ttttcacgtggcagcgagatcacacctca----- 3184

Db 65967 TTTTCACTGGCAGCCAGGATCATCACCCCTCAAGGTAGGCATGAACATTTTATTAATCAAA 66026
QY 3185 ----- agtgtatgtccccaataagtcgagatgttgagta 3218
Db 66027 CTTTCCCTTACTTAACGTGCTCCATAGGTGATTGCCCTCAATAAGTCGGATGTTGAGTA 66086
QY 3219 tacgtctccacggtcactgtggccctgtgaacccctgtctcttggatcgcgtggcaacccgtg 3278
Db 66087 TAGCTCCACGGCTACTGTGGCCGTGAACCTGTCTCTTTGGTGGATCGTGCACACCTGG 66146
QY 3279 cacaggggggtctgggtcccgagagcgctgtctgtcgtatggatctgttccacgggagc 3338
Db 66147 CACAGGGGGGTCTGGGTCCCGAGCAGCGCCTGCTCTCGGTATGGGATCTCTTACAGGGGAG 66206
QY 3339 ctgcattgataataatacaagctcagacgagagccgtcaactccctggccgtgctggcccaag 3398
Db 66207 CTGCATGTATAATATACAAGCTCACAGCAGGCGCGTCACTGCTGGCTGTGGCCCGCAG 66266
QY 3399 ttacgttaactcgtcgttaggcagcagagagaggttgcgtatgggaacgatttccagggaaa 3458
Db 66267 TTACGTAATCTCGCTAGGCACGAGAGAGGATTGCGTATGGGAACGATTTTACGGGAAA 66326
QY 3459 cctgttgactaccatacaacatctcaaacgcatactcagcctactgctgaacaccgtc 3518
Db 66327 CCGTGTGACTACCATCAACATCTCAACAGCATACTCGAGCCTACTGATGCTAACACCGCTC 66386
QY 3519 actattggttacagcaaaatggtaagccctcattcttggatgccaataataagaggac 3578
Db 66387 ACTATTGGTTACGAGCAAAATGGTAAGCCCTCACTCTTGTATTGCCAATATAGAGGGAC 66446
QY 3579 agtaataataaatttaattccaacacagagatctcttattgtggtgagtgctgcactgg 3638
Db 66447 AGTAATAATAAATTTAATTCACACACAGGATCTCTTATTGTGGGATGTGGCAGCTGG 66506
QY 3639 gaagcggctcgcgaagtcaaaactggacttgcgaacccctgcgctctgtcccaaaataat 3698
Db 66507 GCAGCCGGCTCGCAGGTCAAACTGGACTTTTGCAAACTCTGCAAGCTGTGCCAAATAAT 66566
QY 3699 gatcgtcgtcgtcgtcgttaggttgacactacaggaatagatccgctcgtcgcctt 3758
Db 66567 GATGCTTGCTCGGATTCGGTAGTTTGGCAGCTACGGAATGAGATCCGCGTGTCCGCTT 66626
QY 3759 tccatcgtggcagacaagtgcattaaagcgcgaataattttatttagcgtggttcgcta 3818
Db 66627 TCCTATCGTGGCAGACAAGTGCATTAAAGCGCAAAATTTTAAATTTAGCGTGTTCGCTA 66686
QY 3819 gaactaggaataagttgacttaagccttaaaacgcctggaagtcattgacgcattcac 3878
Db 66687 GCACCTTAGGAATAAGTTGACTTAAGGCTTTAAACCGCCTGGAAGTCAATTCACGCCATTAC 66746
QY 3879 tatattataataataatacatattaggttcgcgagcaactacaggttttaacacaagc 3938
Db 66747 TATTTTATATAATATATACATTTATAGGTTCGGCAGCAACTTACGGTTTTTACACAAGC 66806
QY 3939 tgtacgtatctcatctctcagaattttgtttagtttggacactaaagtgttaacagctac 3998
Db 66807 TGTACGTATCTCATCTCTAGAAATTTTGTGTAGTTTGTGACACTAAAGTGTAAACAGCTAC 66866
QY 3999 gctccggtaggttaaggaactaaactaaatgaatcagataataacacatatattttcgcc 4058
Db 66867 GCTCCGGTAGGTTAAGGAACCTAAACTAAATGAATCAATGATATATACACATATATTTTCGG 66926
QY 4059 taattataaaactacatagtcgtcttaagcgcctcagcctcactaataataaactgactaaatg 4118
Db 66927 TAATTATATAACTACATAGTGTCTTTAAAGCGCCTCAGCCTTAATAAATGACTAAATG 66986
QY 4119 ttaaaataaa 4128
Db 66987 TTAATAATAAA 66996
RESULT 2
AC007121

LOCUS AC007121 135182 bp DNA INV 20-MAR-1999
DEFINITION Drosophila melanogaster, Chromosome 2R, region 42A8-42A16, pl clones DS06954 and DS05325, complete sequence.
ACCESSION AC007121 AC005424 AC005426
VERSION AC007121.1 GI:4454581
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 135182)
AUTHORS Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Sequencing of Drosophila chromosome 2R, region 42A8-42A16
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 135182)
AUTHORS Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-1999) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
COMMENT On Mar 20, 1999 this sequence version replaced gi:4028926 gi:4079609.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu
PI library locations: 73-42, 56-45.
FEATURES
source
1..135182
/organism="Drosophila melanogaster"
/strain="y2; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="42A8-42A16"
/clone_lib="PI library, partial Sau3A in pAd10sacBII"
/clone="Pis DS06954 (D338) and DS05325 (D340)"
/note="DS06954 (D338) and DS05325 (D340) were submitted as a project. DS05325 extends from pl end at bpl to pl end at bp 77,598. DS06954 extends from pl end at bp 67,666 to pl end at bp135,182."
BASE COUNT 39447 a 28166 c 28329 g 39240 t
ORIGIN
Query Match 79.4%; Score 3279.6; DB 34; Length 135182;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 3423; Conservative 0; Mismatches 9; Indels 118; Gaps 2;
QY 697 acgatatttcagataacataacctaagatttcgcgcgcggaatgctgttt 756
Db 131437 ATGTTGTCCATTTTCAGAACCTACAGAATCAAAAGTTTCGCGCGGAATGCTGTTT 131496

2 (bases 1 to 313072)
Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

location/Qualifiers
1. 313072
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
complement(join(1392..1813,2404..2648,2713..3050,
3303..4720,4806..4989,5080..5582,5806..6319,6640..7246,
13906..14036,14558..14629))
/gene="CG12110"
/product="CT40213"
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
complement(join(1392..1813,2404..2648,2713..3050,
3303..4720,4806..4989,5080..5582,5806..6319,6640..7246,
13939..14127,14558..14634))
/gene="CG12110"
/product="CT40211"
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
complement(join(1712..1813,2404..2648,2713..3050,
3303..4720,4806..4989,5080..5582,5806..6319,6640..7139))
/gene="CG12110"
/note="CG12110 gene product [alt 1]"
/codon_start=1
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
/protein_id="AAF57264.1"
/db_xref="GI:7302168"
/translation="MADAGTVGGALPPYALTNSEDEHSGSEEDASEENSENNEEGGVF
RDCDEAVVEHNRLCPDFQSLVDSYDETLPFDSVTILSNVGDKPVLVERKETDD
DEEFDDEENSVVLRHELPFTSIYGPSVKFNSQKRVFIPGREIHVRIYDTERSVIT
HLLNPNIYIELTHGPKFKWTIKRKYKHFNSLHQOLSFFRTSLNIPFPFSRSHKERTTL
KATAREMADESTLKDLPSTKVKQTSITPLRAEGRSSKIAGSNANNAMAMISPNHSSIL
AGLTPRRIOKKRKKKKRKLPRFPNRPESLVTVENLSVRIKOLEDYLYNLNLSLYRS
HETLNFVEVSNVSPVCMGILKKEGVILKPTGSTRPGQACNFPFCQKCCVPCNY
FCSDVVCCTWRNRFVVKETCFVIRPTDGSIRAVILFDQGFDSVSTGIOTGMKGLQ
VLTNNRHIVLKCWTRRCKEMOYLKNTANSYARDFTLNPHMSFAPRANTHATCPE
IYMRPALDGDYWRLLDKILLRKAQGVRFVLLYKEVEMALGINSYKSLAKHENI
KVMRHPDHARGGILLWAHHEKIVVIDOTYAFMGGIDLCYGRWDDHHRLTDLGSLTS
SFGSSTRTPSLYFTKDDTDSAFGRSKSSRNAHYDTSAKERPPSPDPENPTSIELKT
LKPDGRLLIPSTLVSSPGETPAESGIALEGMKLNTPEMERKNVLDRLKNNAMKGARMG
KDFMRLTATETEKSAEVYITIESEATDEHVNLMASGGQEVAITTSSQILSEFCG
OAKYWGKDYSNFILKDMNLSNPVDIIDRTTPRMPHWDVGLCVGTSARDVARHP
IORNAMKLEKLRNTRFPIPMKSYHOVRLNPNIQONQORVTCQLLRSVSAWSCGF
LEADLVEOSHIDAYIQITTKAQHYVYIENOFFITMOLGMGVPGAYNNRNOIGETLFK
RIVRAHYETKILILADLMFSSSRKRPFRVYIMPLLPFGEDGVGGSTGIAVRAI
THWYASISRGRTSILTRLOEAGIANPENYISFSLRNHSLFNNTPIETELIYVHSKLL
IADRVVICGSANIDRSMIGKRDSEIAAILMDEFEDEGRMNGKKYPSPGFAGRLRY
LFKPHGLLSEGSRRDLDINDPVECFWHGTWRHISMONTETIYDEVFKCIPDFVK
TFASLRKYOEPPPIAKTAPDLAANRADIOGLVDLPLEFLNKEVLTTPPGTSKEGLIP
TSVWT"
complement(join(1712..1813,2404..2648,2713..3050,
3303..4720,4806..4989,5080..5582,5806..6319,6640..7139))
/gene="CG12110"
/note="CG12110 gene product [alt 3]"
/codon_start=1
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
/protein_id="AAF57265.1"
/db_xref="GI:7302169"
/translation="MADAGTVGGALPPYALTNSEDEHSGSEEDASEENSENNEEGGVF
RDCDEAVVEHNRLCPDFQSLVDSYDETLPFDSVTILSNVGDKPVLVERKETDD
DEEFDDEENSVVLRHEIPFTSIYGPSVKFNSQKRVFIPGREIHVRIYDTERSVIT
HLLNPNIYIELTHGPKFKWTIKRKYKHFNSLHQOLSFFRTSLNIPFPFSRSHKERTTL

KATAREMADESTLKDLPSTKVKQTSITPLRAEGRSSKIAGSNANNAMAMISPNHSSIL
AGLTPRRIOKKRKKKKRKLPRFPNRPESLVTVENLSVRIKOLEDYLYNLNLSLYRS
HETLNFVEVSNVSPVCMGILKKEGVILKPTGSTRPGQACNFPFCQKCCVPCNY
FCSDVVCCTWRNRFVVKETCFVIRPTDGSIRAVILFDQGFDSVSTGIOTGMKGLQ
VLTNNRHIVLKCWTRRCKEMOYLKNTANSYARDFTLNPHMSFAPRANTHATCPE
IYMRPALDGDYWRLLDKILLRKAQGVRFVLLYKEVEMALGINSYKSLAKHENI
KVMRHPDHARGGILLWAHHEKIVVIDOTYAFMGGIDLCYGRWDDHHRLTDLGSLTS
SFGSSTRTPSLYFTKDDTDSAFGRSKSSRNAHYDTSAKERPPSPDPENPTSIELKT
LKPDGRLLIPSTLVSSPGETPAESGIALEGMKLNTPEMERKNVLDRLKNNAMKGARMG
KDFMRLTATETEKSAEVYITIESEATDEHVNLMASGGQEVAITTSSQILSEFCG
OAKYWGKDYSNFILKDMNLSNPVDIIDRTTPRMPHWDVGLCVGTSARDVARHP
IORNAMKLEKLRNTRFPIPMKSYHOVRLNPNIQONQORVTCQLLRSVSAWSCGF
LEADLVEOSHIDAYIQITTKAQHYVYIENOFFITMOLGMGVPGAYNNRNOIGETLFK
RIVRAHYETKILILADLMFSSSRKRPFRVYIMPLLPFGEDGVGGSTGIAVRAI
THWYASISRGRTSILTRLOEAGIANPENYISFSLRNHSLFNNTPIETELIYVHSKLL
IADRVVICGSANIDRSMIGKRDSEIAAILMDEFEDEGRMNGKKYPSPGFAGRLRY
LFKPHGLLSEGSRRDLDINDPVECFWHGTWRHISMONTETIYDEVFKCIPDFVK
TFASLRKYOEPPPIAKTAPDLAANRADIOGLVDLPLEFLNKEVLTTPPGTSKEGLIP
TSVWT"
complement(join(2380..2540,6640..7246,13939..14624))
/gene="CG12110"
/product="CT6556"
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
complement(join(2519..2540,6640..7139))
/gene="CG12110"
/note="CG12110 gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0012110"
/db_xref="FLYBASE:FBgn0033075"
/protein_id="AAF57266.1"
/db_xref="GI:7302170"
/translation="MADAGTVGGALPPYALTNSEDEHSGSEEDASEENSENNEEGGVF
RDCDEAVVEHNRLCPDFQSLVDSYDETLPFDSVTILSNVGDKPVLVERKETDD
DEEFDDEENSVVLRHEIPFTSIYGPSVKFNSQKRVFIPGREIHVRIYDTERSVIT
HLLNPNIYIELTHGPKFKWTIKRKYKHFNSLHQOLSFFRTSLNIPFPFSRSHKERTTL
KATAREMADESTLKDLPSTKVKQTSITPLRAEGRSSKIAGSNANNAMAMISPNHSSIL
AGLTPRRIOKKRKKKKRKLPRFPNRPESLVTVENLSVRIKOLEDYLYNLNLSLYRS
HETLNFVEVSNVSPVCMGILKKEGVILKPTGSTRPGQACNFPFCQKCCVPCNY
FCSDVVCCTWRNRFVVKETCFVIRPTDGSIRAVILFDQGFDSVSTGIOTGMKGLQ
VLTNNRHIVLKCWTRRCKEMOYLKNTANSYARDFTLNPHMSFAPRANTHATCPE
IYMRPALDGDYWRLLDKILLRKAQGVRFVLLYKEVEMALGINSYKSLAKHENI
KVMRHPDHARGGILLWAHHEKIVVIDOTYAFMGGIDLCYGRWDDHHRLTDLGSLTS
SFGSSTRTPSLYFTKDDTDSAFGRSKSSRNAHYDTSAKERPPSPDPENPTSIELKT
LKPDGRLLIPSTLVSSPGETPAESGIALEGMKLNTPEMERKNVLDRLKNNAMKGARMG
KDFMRLTATETEKSAEVYITIESEATDEHVNLMASGGQEVAITTSSQILSEFCG
OAKYWGKDYSNFILKDMNLSNPVDIIDRTTPRMPHWDVGLCVGTSARDVARHP
IORNAMKLEKLRNTRFPIPMKSYHOVRLNPNIQONQORVTCQLLRSVSAWSCGF
LEADLVEOSHIDAYIQITTKAQHYVYIENOFFITMOLGMGVPGAYNNRNOIGETLFK
RIVRAHYETKILILADLMFSSSRKRPFRVYIMPLLPFGEDGVGGSTGIAVRAI
THWYASISRGRTSILTRLOEAGIANPENYISFSLRNHSLFNNTPIETELIYVHSKLL
IADRVVICGSANIDRSMIGKRDSEIAAILMDEFEDEGRMNGKKYPSPGFAGRLRY
LFKPHGLLSEGSRRDLDINDPVECFWHGTWRHISMONTETIYDEVFKCIPDFVK
TFASLRKYOEPPPIAKTAPDLAANRADIOGLVDLPLEFLNKEVLTTPPGTSKEGLIP
TSVWT"
complement(join(16896..17675,17748..17894,17986..19055,
19117..19315,20599..21889,22351..22643,22718..23392,
35382..35595))
/gene="CG8276"
/note="Transcript structure is suspect."
/product="CT24495"
/db_xref="FLYBASE:FBan0008276"
/db_xref="FLYBASE:FBgn00033073"
complement(<16896..>35595)
/gene="CG8276"
/db_xref="FLYBASE:FBan0008276"
/db_xref="FLYBASE:FBgn00033073"
complement(join(22552..22643,22718..22967))
/gene="CG8276"
/note="CG8276 gene product; Transcript structure is
suspect"
/codon_start=1
/db_xref="FLYBASE:FBan0008276"
/db_xref="FLYBASE:FBgn00033073"
/protein_id="AAF57267.1"
/db_xref="GI:7302171"
/translation="MEXRLSDSPGDCRVTSTMTPTLRLDQTSROEPLPOPDNCPAA
APGSKSPTPLPVGKSGQAQHQRAPQOQGGKRNKNAWIYGLLGHSDSPSHCGVGS
CGGRHPEYP"
complement(<29911..>30933)
/gene="CG8330"
/product="CT4956"
/db_xref="FLYBASE:FBan0008330"
/db_xref="FLYBASE:FBgn00033074"
complement(<29911..>30933)
/gene="CG8330"
/db_xref="FLYBASE:FBan0008330"
/db_xref="FLYBASE:FBgn00033074"
complement(29911..30933)
/gene="CG8330"
/note="CG8330 gene product"
/codon_start=1

Oy 2797 cctgcgggattcttaagtgtcataccagggaagcgcgaaagaaacccagggtataacc 2856
Db 228781 CTGCGGGATTGCTTAAAGTGTCCATACCAGGAAGACGCCGAACAGAAACCAGGGTATAACC 228840
Oy 2857 cacatccacctgaacgcgagatcgagtgattgtggcgctcttaatggcgactagatttt 2916
Db 228841 CACATCCCACTCAAGCGGCATCGAGTGATTGTGGCGCGCTCTTAATGGCGGACTAGATTTT 228900
Oy 2917 tacccttagagactactacaagggaagcaaatcgactgggtttttacctcgcttac 2976
Db 228901 TACCCTTTAGACCGCTACTACAAAGGGGAAGCAATCGACTGGGGTTTACCTCGGCTTAC 228960
Oy 2977 agagaaa----- 2983
Db 228961 AGGAGAAGTATGCTTTTCGTTTAAATACCTTTTAAAAAACCTCTAAAGAAAGACACTAT 229020
Oy 2984 -----ctcattgtcgaactgatccactggaagcctgggataatgttgcgagcg 3034
Db 229021 TTCTTTCAGCTCATGTTTCCAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGC 229080
Oy 3035 agcgcgtgcagcaagaacatcccagagaagaccaccaagagagaaatgaaaatcacattgg 3094
Db 229081 AGCGTGTGACGAAGAAGCATCCCAAGACACCACCAAGAGGAATGAAATTCACATTGG 229140
Oy 3095 aggggtgaagactagcccatcagcagcgaatcacatgatgcaggtcgtttaaagacatgg 3154
Db 229141 AGGGTGTAAAGACTAGCCCATCAGCAGGCAATCACATGTCATGCAGTCTGTTAACGACATGG 229200
Oy 3155 ttttcactggcagcgagatcacacccctca----- 3184
Db 229201 TTTTCACTGGCAGCGAGGATCACACCCCTCAAGGTAGGCATGAACCATTTTATTAATCAA 229260
Oy 3185 -----aagtgtattgcctcaataaagtctgagtggttgagta 3218
Db 229261 CTTTCCTTACTCTAACTGGTCTCCATAGTGTATTGGCTTCAATAAGTCGGATTGAGTA 229320
Oy 3219 tacgctcacagtgactggtggcctgtgaacctgtctcttgtggatcgtcggaacctgg 3278
Db 229321 TAGCTCCACGGTCACTGTGGGCCCTGTAAACCTGTCTCTTTGTGGATCGCTGCAACCTGG 229380
Oy 3279 cacaggggggtctgggtcccaggaagcgcctcgtctgcgtatgggtatcgttcaacggggagc 3338
Db 229381 CACAGGGGGGTCTGGGTCCCAGGACGGCTGCTCTGCGTATGGGATCTGTTCAACGGGAGC 229440
Oy 3339 ctgatgtataatacacagctccacgagcgagcctcagctcagctgcctggccttgcccccag 3398
Db 229441 CTGCAATGATTAATATACAAGCTTCAACGAGCGGCTGAGTCCCTGAGTCCCTGFGCCGCCAG 229500
Oy 3399 ttacgtaactcgtctaggcagcgagagagatttgcgtatgggaacgatttcaggggaaa 3458
Db 229501 TTACGTAATCTCGTTAGGCACGGACGAGAGATTTCGCTATGGGAACGATTTTCAGGGAAA 229560
Oy 3459 cctgttgactaccatcaacatctcaacgcgacatactcgcgactctatgcttaaacccgtc 3518
Db 229561 CCGTGTGACTACCATCAACATCTCAAGCGCATACPCGAGCGCTACTGATGCTTAACACCCGTC 229620
Oy 3519 actattggttacgaacaaaatgggtgaagcctcattcttgattggccaataaagaggac 3578
Db 229621 ACTATTGTTAGGACAAATGGGTAAAGCGCTCAATTCCTTGATTGCAATATAAGAGGGAC 229680
Oy 3579 agtaataataaatttaattccaacacagagatctcttattgtgtgggtatgctgcgactgg 3638
Db 229681 ACTAAATAATAATTAATTCACACAGGATCTCTTATTGTGCGGATGTCCGCACTGG 229740
Oy 3639 gcagccggctcgcgaggttcaaaactggactttgcaaacctcgcagctctctccccaaaataat 3698
Db 229741 GCAGCGGCTCGCGAGGTCAAACTGGACTTTGCAAAACCTGCGAGCTCTGCCCAAAATAAT 229800
Oy 3699 gatctctccctcgcgattcggtagtttgactacaggaataagatccgcgctgcgcgtt 3758
Db 229801 GATGTTCCCTGCGGATTCGGTAGTTTGGGACTTACGGAATGAGATCCCGCTCGTCCGCTT 229860
Oy 3759 tccatcgtgcgacaaagtgccattaaagcgcaaaatttttaatttagcgtgggttcgcta 3818

Db 229861 TCCTATCGTGGCAGCAAGTGCCATTAAAGCGCAAAATTTTAAATTAGCGTGTGTCGCTA 229920
Oy 3819 gcacataggaaataattgacttaaggctttaaacgacctggaagtcatattgaacgactaac 3878
Db 229921 GCACCTAGCAATTAAGTTGACTTTAAGCTTTAAAGCCCTGGAAGTCATTGACGCAATTCAC 229980
Oy 3879 tattttataataatatacacattatagggtccgcagacaactacggtttttaacacaagc 3938
Db 229981 TATTTTATATAAATATATACACTATTAGGGTCCGCAGCAACTTTACGGTTTAAACAACAGC 230040
Oy 3939 tgcagtatctcatctctagaatttttgtttagttgttgacacactaaagtgtacaagcttac 3998
Db 230041 TGTACGTATCTCATCTCTAGAAATTTTGTGTAGTTTGTGGACACTAAAGTGTAAACGCTAC 230100
Oy 3999 gctcggtaggttaaggaaactaaaalgaatcagatatatacacatatattttcgcg 4058
Db 230101 GCTCCGGTAGGTTAAGGAACCTAACTAAATGAATGATATATACACATATATTTTCGGC 230160
Oy 4059 taattatataaactacatagtgctttaaagcgcctcagcctaataataaactaactaaatg 4118
Db 230161 TAATTATATAAACTKATAGTGTCTTAAAGCGCCTCAGCCTAATATAAAATGACTAAATG 230220
Oy 4119 ttaaaataaaa 4128
Db 230221 TTAAAAATAAA 230230
RESULT 4
AC020090 60625 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC020090
VERSION AC020090.1 GI:6664807
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 60625)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212008 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..60625
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 17399 a 13001 c 12988 g 17237 t
ORIGIN
Query Match 57.9%; Score 2390.4; DB 43; length 60625;
Best Local Similarity 97.18; Pred No. 0;
Matches 2480; Conservative 0; Mismatches 11; Indels 53; Gaps 2;
Oy 697 acgatattttccattacataacctacagaaatacaaaagtttccgcgcgcggaatatgctgttt 756
Db 57951 ATGTTGTGTCATTTTCAGAAACCTACAGAAATCAAAAGTTTCGCCGCGGCAAAATGCTGTTT 58010
Oy 757 ggaattaccatcgacgagacacttgattcaagcgctctccattcgcgctcggtcgcgattt 816
Db 58011 GGATTTACCCATGACAGGACACTGGATTTCAGGCGCTATCCATTGCGCGCTCGGTCGCGTAT 58070
Oy 817 atacagtagccttgacgttattctctcaagcacacgatatggatgtaictggacactcta 876

||||| Db 58071 ATACAGTATG-CCTGAGCTTATTCCTCAAGCACACAGATATGAGTATCTGGACACTCTA 58129
Qy 877 aagaaaaagctgctgcacactatccccactccgttggttagtcgctgaagag 936
Db 58130 AAGGAAAAGCTGTGGGACACTATCCCCACTCCCTGGCTAGTCGGCTGAAGAG 58189
Qy 937 ccaagacataaacttacatcttttatccagagagatcacaggatgaggagctggcgct 996
Db 58190 CCGACACCATATCTTATCTTTTATCCAGGAGATACAGGATGCGGAGCTGGTGCCT 58249
Qy 997 tacacagtgcccttatgttggttggttatgttactctctctgttcgaaaaatcgat 1056
Db 58250 TACACAGTGGCCCTTATGTGTGGTGTGTGTATGTACTTCTCTGTTCGAAAAATCGAT 58309
Qy 1057 gtatttcgtcccgctttttgtgcttagcttatgttagcgtataaccacacccgggagcttg 1116
Db 58310 GTATTTCTGTTCCCGCTTTTGTCTGGCTTATGTAGGCTAATCAACACGCCGGAGCTTG 58369
Qy 1117 gccatgccccttggtgtgttcttcttcttgccctgacaaattcgtcgactcaaaaggac 1176
Db 58370 GCCATGTCCTTGCCCTGTGTCTTTCTTTTGGCTGACAAATTTGCTGAGTCAAGGAC 58429
Qy 1177 attttccctacactgtaactcccttggtggttggaataagcttggtgatcacaaagagc 1236
Db 58430 ATTTTCCCTACCTTGTAACTCCTTGTGGATGTGAAATAGCTTGGTGATCACAAGAGC 58489
Qy 1237 gtagtccaatgaagacacattcgagtgaaagatccgctgagcagctcttagcaag 1296
Db 58490 GTAGTCTCAATGACGAGACATTCGACGTGAAGATCCGGGTGGCGAGGCTCTTAGCAAG 58549
Qy 1297 gaggttggtcatataccaagactctttgaagagatacaaatcttgacaaattggtctt 1356
Db 58550 GAGGTTGGCATATATCCAGACTCTTTTGACGGAGATACAAATTTTGACAATTTGCTCTT 58609
Qy 1357 gctacttcgtgcccctgcactccaggagtttgttatctttgccaatagtcggttgccttcc 1416
Db 58610 GCTACTTCTGTGCCCCGTATCCAGGAGTTTGTATCTTTGCCATATGTCGGCTTGTCTTC 58669
Qy 1417 gatttatgtacagatgctgctctctctcaaaactcgccatgaacattaaagcgacc 1476
Db 58670 GATTTATGCTACAGATGTGCTCTACTCAACAATAGTGGCCATGAACATTAAGCGGACC 58729
Qy 1477 gagtaacgaggagggcgaagcactctctaagatgtgctgagctgcaccaagggct 1536
Db 58730 GAGTATACGGCGGAGGCCAGCACCCTCCTAAGATGTTGCTGAGCTGCACCCAGGGGCT 58789
Qy 1537 ggtcgacaggaattccgatttttcggggcgcccgacactgccaacgtttgtccctggc 1596
Db 58790 GGTTCGACAGGATTCGGATTTTTCGGGGCGCCGCCAGCACTGCGACCGTTGTCCCTGGC 58849
Qy 1597 acatttcagcgttctcagtcgcatacaaaactgtgttttgctgactccgcactctgttagc 1656
Db 58850 ACATTTACGGTTCTCAGTCGCATCCAAACTGTGTTTGTCTATCCCGATCTGTTAGC 58909
Qy 1657 gatcgtacaagcttggttaatgacactcgtgcggagcaacgaatacccaaacgcata 1716
Db 58910 GATCGTACAAGCTTGTTAATGACACTCGTCGCGGAGCAACGAATACCCAAACGCATA 58969
Qy 1717 aagattgtaaatcttggtgcggcggaactcgttttttcagcgtgccttcgatctgtgagc 1776
Db 58970 AAGATTGTAAATTTCTTGGCGCGGACTCGCTTTTTCAGCGTGCCTTTCATGATCTGGATG 59029
Qy 1777 atttgtgatgatgctctatgtttataattcgggatatctgagcagagtttttagcatg 1836
Db 59030 ATTGTGTGATATGCTCTCTAGTATTTATAATTCGGGATATCTGGAGCAGTTGTTAGCATG 59089
Qy 1837 cagagcaagcgacaatgacggcaaccccttgaactcaacggcgactacagcggtgctg 1896
Db 59090 CAGAGCAACGGCACATGACGGCAACCCCTGAACCTTCAACGGGACTACAGGGGCTCGG 59149
Qy 1897 ggagcagtcagcagtttttctcaggggatggcaagcgagcgtgccacagtgctg 1956
|||||

Db 59150 GGAGCAGTCAGCAGCTTTTTCGAGGGATGGCAAGCGGAGCGGTGCCACAGTGGC 59209
Qy 1957 ccaagcgaagcgcttttctacgcgaataaaaagctccctctagcgtacatataaacgaa 2016
Db 59210 CCAAGCGGAAGCGCTTTTCTACGCCAATAAAGCTCCCTAGCGATCGATATAAACGAA 59269
Qy 2017 agggcgaagaaatgatgagacttgatataccagcttcgacctaataactatttccttca 2076
Db 59270 ACGGCGAGGAATGATGAGACTTCGATATCCAGCTTCGACCTAACTATTTTCCCTTCA 59329
Qy 2077 aaactccactggtccacgattatgaacagtgataacacatctcactaagtgggcactagtt 2136
Db 59330 AACTTCCACTGGTCCAGATTTATGAACAGTACAACTCTCACTAAGTGGCCTACGTT 59389
Qy 2137 accctgctaccacacattcccttagctatgcactgcctccgagctagcactctgtg 2196
Db 59390 ACCCTCTACCGACCATTCGCCTTAGTCATGCCATCGCTCCGAGGTAGCACCTCTGTTG 59449
Qy 2197 cggaaatcccgagagcagctgcaacaaaattttcaatggaagcccttagcgtgcactc 2256
Db 59450 CGGAATCCGACGAGCAGCTGCAACAAAATTTTCAATGGAAGCCCTAGCCGCTGCATC 59509
Qy 2257 gatccgctggaacttaacgatgaacgctgcccgtgagctccgctcgtatgtaatggcagag 2316
Db 59510 GATCCGCTGGACTTTAACGATGACGAGCTGCCCGCTGAGTCTCCGATGGTAAATGGAGAG 59569
Qy 2317 gggttgctctggttcccaagagcccatggaataatttttgcacatctctctgtgctgc 2376
Db 59570 GGGTTGCCCTCTGTTCCCAAGACCCCATGGAATATTTTTCGCCATCTCTTGTGCTGC 59629
Qy 2377 atcagcatcttctgctttgtcacagatggtggttttctacccgtgcataatgacag 2436
Db 59630 ATCAGCATCTTCGTGCTTTGCTACAGATGGTGGTTTTCACCGCTGCATATGTACCAGG 59689
Qy 2437 aactatgcccagtgggcgtccagttggcacgaactctgagcccgataagcagactgag 2496
Db 59690 AACTATGCGAGTGGCGCTCCAGTTGGCACGAATCTGAGCGACCGTACAAGCAGACTGAG 59749
Qy 2497 caaatctgagaggagttccaaagcaaatccgcgacacacaaatcgacttaaatccctg 2556
Db 59750 CAAATCTGGAGGAGTTCCACGCAAAATCGCCGGACACAAACATCGATTGAATGCCCTG 59809
Qy 2557 gtctctgacggcgctacataatcagctgctgctttaaaggccaaatcccgagtgtgggat 2616
Db 59810 GTGCTCTGCGCGCTTACATAATACAGCTGCTGCTTAAAGGCCAAATCCGAGTGGGAT 59869
Qy 2617 gaacgagtgggagcagctaacacgcatctcccgaiccgatattcagatctctcagcag 2676
Db 59870 GCACGAGTGGCGAGCAGCTAACACAGCATCTCCCGATCCGATATTCAGATCTCTCAGCAG 59929
Qy 2677 cggacggtggcagacgctggtacgaaagctggccgtgctcacccgtctggtgcttgac 2736
Db 59930 CGGACGATGGCGAGCAGCTGTTGACAAAGCTGGCGCTGTACACCGCTCTGGTCCCTTGAC 59989
Qy 2737 tacttcgataatctaatcgcagtagctgcgcaacgycgcggtgaaattgtgggaaatcc 2796
Db 59990 TACTTCGATAATCTAATTCGACGTAGGCTGGCCACAGCGCGGTAGAAATTTGGGAATCC 60049
Qy 2797 cctgvggagattcttaagtgtgcataccaggaagacgcgaagaaacagggtataaac 2856
Db 60050 CTGCGGGATTCCTTAAGTGTGCATACAGGAAGACGCGGAAGAAACAGGGTATTAAC 60109
Qy 2857 cacatccactgaacggcgatcgagtgattggcgcgcttctaaatggcgactagattt 2916
Db 60110 CACATCCACTGAACGGCGATCGAGTGATTGTGGCGCTCTTAATGGCCGACTAGATTTT 60169
Qy 2917 taccgcttagacgctactacaaggggaagcaaatcgactgggggttttaccctcggttac 2976
Db 60170 TACCGCTTAGACAGCTACTACAAGGGGAAGCAATCGACTGGGGTTTTACCTCGGCTTAC 60229
Qy 2977 agggaaa----- 2983
Db 60230 AGGAGAAGTATGCTTTCCTTTAAATACCCCTTTTAAAAAACTCCTTAAAAAGAAAGACTAT 60289

```
QY 2984 -----ctcatgttcgaactggtatccactggaagcctgggattaaattgttgcagcgc 3034
Db 60290 TTCTTCACGCTCATGTTCAAACTGGATCCACTTGAAGCCTGGGATTAAATGTTCCAGCAC 60349
QY 3035 agcgtgtcagcaagaagcatcccagaagaccaccaagagagaaatgaaatcacattgg 3094
Db 60350 AGCGGTGTTCAGCAAGAAGCATCCAGTAGACACCACCAAGGAGAAATGAAATCACATTGG 60409
QY 3095 aggggttaagactagccaatcagcagccaatcacatgcattgcaggtcgtttaacgacatgg 3154
Db 60410 AGGGTGTAAAGCTAGCCCATCAGCAGCCAAATCACATGCGATGCGAGTTCGTTAAACGACATGG 60469
QY 3155 ttttcactgagcagcaggtatcacacctcaaggt 3188
Db 60470 TTTTCACTGGCAGCCAGGATCACACCTCAAGGT 60503

RESULT 5
AC014229
LOCUS AC014229 12978 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014229
VERSION AC014229.1 GI:6437106
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Adams M. and Venter J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212393 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..12978
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3902 a 2637 c 2629 g 3810 t
ORIGIN
Query Match 27.3%; Score 1126.4; DB 42; Length 12978;
Best Local Similarity 91.4%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 1; Indels 118; Gaps 2;
QY 2863 caccctgaacggcgatcgagtgattgtggcgctcttaattggccgactagattttaccgc 2922
Db 11 CCCCTGAACGGCGATCGACTGATTGTGGCCGCTCTTAATGGCCGACTAGATTTTACCOC 70
QY 2923 tttagaagctactacaaggggagcaaatcgactggggttttacctcggttacaggaga 2982
Db 71 TTAGAGACGTACTACAAGGGGAAGCAAAATCGACTGGGGTTTACCTCGCTTACAGGAGA 130
QY 2983 a----- 2983
Db 131 AGTATGTCCTTTCGCTTAAATACCCCTTTAAAAAACCTCTCTAAAGAAAAAGACTATTTTCCTT 190
QY 2984 ----ctcatgttcgaactggtatccactggaagcctgggattaaattgttgcagcagcact 3040
Db 191 CAGCTCATGTTCCAACTGGATCCACTTGAAGCCTGGGATTAATGTTTGAGCAGCAGCCGT 250
QY 3041 gtcagcaagaagcatcccagaagaccaccaagagagaaatgaaatcacattggaggggtg 3100
Db 251 GTCAGCAAGAAGCATCCCAGAGAGACCACCAAGAGGAATGAAATCACATTGGAGGGTG 310
```

```
QY 3101 taagactagcccatcagcagccaatcacatgcattgcaggtcgttaacgacatagttttca 3160
Db 311 TAAGACTAGCCCATCAGCAGCCAATCACATGTCATGAGGTCGTTAAACGACATGGTTTTTCA 370
QY 3161 ctggcagccaggtatcacaccctca----- 3184
Db 371 CTGGCAGCCAGGATCACACCCCTCAAGGTAGGCATGAANCACTTTTATTAAATCAACTTTCC 430
QY 3185 -----aggtcattgctcctaataagtcggatgtttaagatatcgt 3224
Db 431 TTACTCTAACTGGCTCTCCATAGGTGTATTTCCTTCAATAAGTCGCATGTTGAGTATACGCT 490
QY 3225 ccacgltcactgtggccctgttaacctgtctctttgtgattcgtcgtcaaaccttgccacagg 3284
Db 491 CCACGGTCACTGTGGGCCCTGTAACTCTCTCTTTGTGGATCGCTGGCAACTGGCACAGG 550
QY 3285 ggggtctgggtcccagggcgtcgtctcgtatgggactgtttcacggagagccctgcatt 3344
Db 551 GGGGTCTGGGTCCCAGGACGGCTGCTCTCCGTATGGGATCTGTTACGCGGAGCCTGCAT 610
QY 3345 gtataatacgaagctcacaagcagcagcgttcagctcgttgcctgtgcgttcgtcccgagttacgt 3404
Db 611 GTATAATATACAAGCTCAGCAGCGGAGCGCTCAGCTGCTTGGCTGTGGCCGCCAGTTACGT 670
QY 3405 aatctcgtctaggcagcagcagagagatttgcgtatagggaacgatcttcagggaacactgtt 3464
Db 671 AATCTCGCTAGGCACGAGGAGAGGATTTGCGTATGGGAACGATTTTCAGGGAACCTGTT 730
QY 3465 gactacatacaacatctcaaacqcatcactcgagcctcactgatcgtcaaacccgtcactatt 3524
Db 731 GACTACCATCAACATCTCAACCGCATCTCGAGCCTACTGTATGATGCTAACACCGTCACTATT 790
QY 3525 ggttcagcagcaaatgggttaaggcctcattcttattgccaatataagaggggacagtaaa 3584
Db 791 GGTTCAGAGCAAAATGGTAAAGCCCTCATTTCTTGATTGGCAATATAAGAGGGACAGTAAA 850
QY 3585 taataaatttaattccaacacagggatctcttattgtgtggatgtcgcactgagcagcc 3644
Db 851 TAATAAATTTAATTCACACACAGGATCTCTATTGTGTGGGATGTCCGCACTGGGAGCC 910
QY 3645 ggctcgcgaggtcaaaactggactttgcaaacctgcagctcttcccaaaaataatgatgt 3704
Db 911 GGCTCCGAGGTCAAACTGGACTTTTGCAAACCTGCAGCTCTGTGCCCAAAATAATGATGCT 970
QY 3705 tgcctgcgattcggtagtttgcgactacggaatgagatccgcgtcgcgtcttccctat 3764
Db 971 TGCCCTGCGATTGCGTAGTTTGGCACTACGGAATGAGATCCGCGTGTCTCGCTTTCCCTAT 1030
QY 3765 cgtgcagacaagtgccattaaagcgcataattttaatttagctggttgcgttagcaact 3824
Db 1031 CGTGGCAGACAGTGCCTTTAAAGCGCAAAATTTTAAATTTAGCGTGGTTCGCTAGCACCT 1090
QY 3825 aggaataaagtgtacttaaggctttaaagcgttgaagtcattgacgcattcactatttt 3884
Db 1091 AGCAATAAGCTTGACTTAAGCCTTTAAACGCGCTGGAAGTCATTGACGCATTCACTATTTT 1150
QY 3885 atataaataatacactattagggtccgcagcaactaacgtttttaaccaagcgtgacg 3944
Db 1151 ATATAAATATATACACTATTAGGTCCCGCAGCAACTTACGGTTTAAACACAAGCTGTACG 1210
QY 3945 tatctcatcttagaatttctgttagtttggcacactaagtgtaacagctacgtccgc 4004
Db 1211 TATCTCATCTCTAGAAATTTTGTAGTTTGTGGACACTTAAGTGTAAACAGCTACGCTCCG 1270
QY 4005 gtagtttaaggaaactaaactaaatgaatcagatatatacacatatattttccgtaattta 4064
Db 1271 GTAGGTTAAGGAACCTAACTAAATGAATCAGATATATACACATATATTTTCGGGTAAATTA 1330
QY 4065 tataaactacatagtgctcttaagcgcctcagcctaataataaaaaagactaaaattgtlaaaa 4124
Db 1331 TATAAATACATAGTGTCTTAAAGCGCCTCAGCCTTAATATAAATGACTAAATGTTTAAAA 1390
```

```

QY 4125 taaa 4128
Db 1391 TAAA 1394

RESULT 6
AC009255
LOCUS
DEFINITION
AC009255 109699 bp DNA HTG 10-DEC-1999
Drosophila melanogaster chromosome 2 clone BACR39E09 (D1043)
RPCI-98 39.E.9 map 42A-42A strain y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 103 unordered pieces.
AC009255
VERSION AC009255.3 GI:6554247
KEYWORDS
SOURCE HTG; PHASE1.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 109699)
AUTHORS
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkley,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 109699)
AUTHORS
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 10, 1999 this sequence version replaced gi:5734580.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgfruitfly@berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 103 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1199: contig of 1199 bp in length
* 1200 1279: gap of unknown length
* 1280 1997: contig of 718 bp in length
* 1998 2077: gap of unknown length
* 2078 2686: contig of 609 bp in length
* 2687 2766: gap of unknown length
* 2767 3839: contig of 1073 bp in length
* 3840 3919: gap of unknown length
* 3920 4568: contig of 649 bp in length
* 4569 4648: gap of unknown length
* 4649 5320: contig of 672 bp in length
* 5321 5400: gap of unknown length
* 5401 6335: contig of 935 bp in length
* 6336 6415: gap of unknown length
* 6416 7390: contig of 975 bp in length
* 7391 7470: gap of unknown length

8350: contig of 880 bp in length
8430: gap of unknown length
9510: contig of 1080 bp in length
9590: gap of unknown length
10436: contig of 846 bp in length
10516: gap of unknown length
11243: contig of 727 bp in length
11244 11323: gap of unknown length
11324 11946: contig of 623 bp in length
12026: gap of unknown length
12701: contig of 675 bp in length
12781: gap of unknown length
13062: contig of 281 bp in length
13142: gap of unknown length
14057: contig of 915 bp in length
14137: gap of unknown length
14805: contig of 668 bp in length
14885: gap of unknown length
15334: contig of 449 bp in length
15335 15414: gap of unknown length
16203: contig of 789 bp in length
16204 17032: contig of 749 bp in length
17112: gap of unknown length
18667: contig of 1555 bp in length
18747: gap of unknown length
19961: contig of 1214 bp in length
20041: gap of unknown length
20778: contig of 737 bp in length
20858: gap of unknown length
21803: contig of 945 bp in length
21884 22739: contig of 856 bp in length
22819: gap of unknown length
23590: contig of 771 bp in length
23670: gap of unknown length
24575: contig of 905 bp in length
24655: gap of unknown length
25422: contig of 767 bp in length
25502: gap of unknown length
26176: contig of 674 bp in length
26256: gap of unknown length
26758: contig of 502 bp in length
26838: gap of unknown length
27932: contig of 1094 bp in length
28012: gap of unknown length
29308: contig of 1296 bp in length
29388: gap of unknown length
30496: contig of 1108 bp in length
30576: gap of unknown length
31728: contig of 1152 bp in length
33080: gap of unknown length
33160: gap of unknown length
34333: contig of 1173 bp in length
34413: gap of unknown length
35478: contig of 1065 bp in length
35479 35558: gap of unknown length
36465: contig of 907 bp in length
36545: gap of unknown length
37699: contig of 1154 bp in length
37779: gap of unknown length
38966: contig of 1187 bp in length
39046: gap of unknown length
40514: contig of 1468 bp in length
40594: gap of unknown length
41980: contig of 1306 bp in length
41980: gap of unknown length
43438: contig of 1458 bp in length
43518: gap of unknown length
44606: contig of 1088 bp in length
44686: gap of unknown length
46717: contig of 2031 bp in length

```



```

RESULT 7
CNS01P6Y 790 bp DNA STS 18-FEB-2000
LOCUS Anopheles gambiae STS SP6 end of clone 27019 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
sequence tagged site.
ACCESSION AL154442
VERSION AL154442.1 GI:7015361
KEYWORDS STS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidea; Culicidae; Anopheles.
1 (bases 1 to 790)
REFERENCE
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 790)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Direct Submission
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
Location/Qualifiers
1..790
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="27019"
/clone_lib="NotreDamel"
/note="end : SP6"
BASE COUNT 159 a 194 c 247 g 178 t 12 others
ORIGIN
source
Query Match 6.4%; Score 263.2; DB 13; Length 790;
Best Local Similarity 64.4%; Pred. No. 1.9e-68;
Matches 382; Conservative 6; Mismatches 205; Indels 0; Gaps 0;

QY 923 cgtcgtggaagcagcagacataacttacatctttttatccaggagatcacaggtgt 982
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 CGGCCCGGAGCAGCAGCATCATGTACATCTCTTACCGGGCGGAATCAACTGGA 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 983 gggagctggcttacacagtgccctttatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 AGAGGTGCTGCCCACTGTATCCGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1043 ttcgaaaaatcgatgtattctcccgctttttgtggtgtgtgtgtgtgtgtgtgtgtgtgt 1102
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 TCCGCAAAATCGAAGCATCCGGTCGGGCATYATGTAGCTCTGTGCGGTGCTGACCG 185
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1103 cagccggagctggccatgtcccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 TGTGGCAGCTTGTCTGATGTATCGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1163 tgcagtaaaaggacattttccctacccttgatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 TCAGTGAAGGTGTCTATCCGTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 305
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1223 tgatcacaagagcgtagtctaatggacagacattcgcgtgaagatccgcgtgtgcgc 1282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 TTATTACGAAGAGTGTCTCACGACGACAAACAATVTGGAGCTWAAGATACGGGTGCGAC 365
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1283 aggcctctaacgaaggaggtgttgcatatccaaagactcttttgacggagatacaaat 1342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 AAGTCTGTAGCCGGGAGGGCTGTGCTCCATTATGAAGAACTTGTCTGACGGAAATTACCATCC 425
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```
/product="SREBP cleavage activating protein"
/protein_id="AAB19103.1"
/db_xref="GI:1675220"
/translation="MTLRLERREKISQAFYNHGLLCASYPIPIILFTGLCLILACYPIL
LKLLPLGPGVFEFTSPVVDSDHKGQEPSEOPWYGVAPVAYIQIFKVS
SPWKNLLAVDFRPLSRAPQLVEEIRHVLRSSTGKSLVEECLVQDILLPGIRKL
RNLLPEHCLLSLPGNFWONDERHADPDIIGTTHOHEPKTLOTSAFLKOLLFGVPG
KYSVSLYTRKRTSYTTLVFORYHAKFLSLRLARMLLHPSNCSLRARLNVHVF
KEEGIAELILVITYIILPAYIFSTRKIDWVSKWGLAALAAVTVYLSLLMSVGLC
TLFGLTFLNGEIEFPYLVVIGLENVLVTKSVSTPVDLEVLKRLIRIAQGLSSWSI
MKNVATLGLIYFTLVPALQEFCLFAVGLVSDFFLQFMFFFTYLSLDIRMELA
DLNKLRLPESCLSAKVPGRPARYQLAVRPAMPHTITLQPSFRNLRLPKRLRVIY
FLAKRLAQLIMAGTVVIGLTVDPAGRLTYLAAQVTEQSPGLSEGLGPMVPVPSG
WLPASRPDPAIFPPDAPKLPNOTVPELPEHAAPAEQVHDSRPAEVTWGPDEEL
VRLSRHPWPLFNWYNTLAKRYLSLPLVPVTLRNPQEALEBROPDGSRSAWAP
ESLPAGLWEAGPKPGGTQAGDITLYKVAALGLAAGLVLLLLCLLRYLCVCPNRYG
PGGAGRRRLGDCDDYGYAPPEIEIPLVLRHLMDECLASDMLLVSCCLAQV
CVMAQDGLCTLRIPRGRSDSCGGAFTQENWRLSDGKTSPEPSPGLRHR
PRGPPQALFGDQDPLTLIDTNFVOLPEPTQPEPRHRAGCGRARDGYDFRLVQ
RVYQEGLAARVMPALRPPSPGSLPOASQEDGAPEKSPPLAWPSTAGSIWSLEL
OGNLIYVGRSSGRLEWMDAIFEVLCSSNDEVSSGITALVFLDRIRIVARLNGSLDFES
LETHLSLPLQFRGTGRGSSPSSSYSSNTVACHLTHVPCAHQKPIALRAAAGR
LVTGSDHHLKRVLEDSCLFLTQGHSGAITTVYIDQTMVLAGSGDGAICLMDVLT
GVSRYHTFAHRGDVTSLTCTTSCVSSISGLDLINWDRSTGIKLYSTOODLGCCASLG
VISDNLVLTGGGCVSEWDLNYGDLLOTVYLKNSQAQPARQIILVLDNAIVCNFSGE
LSLVVPSVLEKLD"
BASE COUNT      824 a 1255 c 1153 g 887 t
ORIGIN
```

```
Query Match      4.5%; Score 187.4; DB 12; Length 4119;
Best Local Similarity 51.5%; Pred. No. 3.7e-45;
Matches 577; Conservative 0; Mismatches 511; Indels 33; Gaps 5;

QY 394 taccgtccagcagatactctgcgaaccagttgtctccctgagcagagaggaatgcagctt 453
Db 393 TACATCCACAGATATTATTGAAGTCATCGGTATCTCCTCGGCACAAAAAATCTCTGGCA 452

QY 454 atggatgccttcgtgcgcgcgtacacgaagtttttaaatcttgtaaatgtgcgcaat 513
Db 453 GTAGATGTGTTCCGATTGCTCTGCCCCGGGCATTCACACTGGTGAGGAATCCGGAT 512

QY 514 caccagagcagcaaaaacaaacgtac-----cctggagcacaactgcctacatgtagac 567
Db 513 CACGTACTGAGAGACAGCTCAGGGACCAAGAGCTGAGGAAGATTTCCTCAGGTGACC 572

QY 568 aacgtataacgcggaacacacacgcgagcgtggaccagatctttccggagatagctgcctg 627
Db 573 GACTTGTGCTCCAGG---ACTCAGGAACCTCCGAAACCTACTTCTCTGAACATGGCTGCCTG 629

QY 628 ctgctctccgcgcgaacctgtggacgcagaactctcagaactttactcgggacacaaac 687
Db 630 CTGCTGCTCCCTGGAAACTTCTGGCAGAAATGACTGGGACGGATTCCATCTGACCTGAC 689

QY 688 atcctgaacacgatatttccag---taccataacctacagaaaataaaagtcttcgcggcg 744
Db 690 ATCATTGGGACTATCCATCAGCATGAGCCAAACAACTCTCAGACATCAGCCACTCAAA 749

QY 745 gaatctgttcttgattaccctatgaggacactgaattcaacgcgtatccattgcgcgt 804
Db 750 GACCTGCTGTTTGGTGTGCTGGGAAGTACAGTGGGGTGTGAGGCTCTA-----CACA 800

QY 805 cagtcgcgtattatcacagtatccttgacgtttattctcacaacacagatggagat 864
Db 801 AGAAGAGACCGGTCTCGTACACCATCACCTCGGTCTTCCACGCGTACCATGCCAAGTTT 860

QY 865 ctggacactctaaaagaaagctgtgcgcacactatccccccactccctgcgttagtcg 924
Db 861 CTGAGCAGCCTACGTGCGCGGCTCATGCTCTCTGACCCCTAGCCCCAACTGCAGCCTTCG 920

QY 925 tcggctgaagaccgcagacacataacttacctttttatccaggagagacagagatgag 984
Db 921 GCCG-----AGAACCTGTGTCACGTGCACCTTCAAAAGAGAGAGATTGGCATGGC 968
```

```
QY 985 gacgtgggtccttacacagtgcccttattgttgggtttgttcttattgttacttctctgtt 1044
Db 969 GAGCTCATCCCCCTCGTGACACCTACATATATCTGTTTGGCTACTACTTCTCCACA 1028

QY 1045 cgaataacgatgtatttcttccgcgcttttgcgtggccttattgtagcgttaacacacaca 1104
Db 1029 CATAAGATCGACATGGTCAAGTCTAAGTGGGACTCGCCCTGGCAGCTGTGGTGCACAGTG 1088

QY 1105 gccgggagcttggccatgtcccttggcttgttcttcttggcctgacaaatttcgctg 1164
Db 1089 CTAGCTCTCTCTCATGTCTGTGGGGCTCTGCACCTCTCTTTGGCCTGACACCCACACTC 1148

QY 1165 cagtcaaaagacatitccccctaccctttaaactcttgggttgggttggaaataagctgggtg 1224
Db 1149 ATGCGGGTGAAGATTTTCCCATACCTGGTGGTGGTTATTGGGTAGAGAAATGTGTGGTG 1208

QY 1225 atcaaaagagcgtagctcacaatgaacgagacattgcagctgaagatccgcgtggcgag 1284
Db 1209 CTCACCAAGTCGGTGTATCAACCCAGTGGACCTCGAGGTGAAGCTTCGGATTGCACAA 1268

QY 1285 gctttagcaagagaggttggcatatataccaagactctttgacggagagatacaaatitg 1344
Db 1269 GGCCTTAAGCAGTGCAGAGCTGGTCCATCAATGAAGAACGCTGGCAACTGAACCTGGCATCATC 1328

QY 1345 acaattgcttgcacttctgcgcgcgtcaccagagtttctgtatcttggccatagtc 1404
Db 1329 CTCATGGCTATTTCACCCCTTGTGCTGTATCCAGGAGTTCTGCTCTTGTGTGTG 1388

QY 1405 ggcttgccttcgattttatgtacagatgctgctcttctcaacaatactggccatgaac 1464
Db 1389 GGCTTGGTCTCTGACTCTCTCTCCAGATGTTCTTTTTCACCACTCTCTCTGTGTCGATCGAC 1448

QY 1465 attagcgacgcagatatacgcgcgagggcgaagcaacacttcc 1505
Db 1449 ATTGCGCAATGGAGCTAGCGGACCTGAAACAAAGCGTCTGCC 1489
```

RESULT 10

```
AC008339 118474 bp DNA HTG 06-AUG-1999
Drosophila melanogaster chromosome 2 clone BACR13P06 (D917) RPCI-98
13.p.6 map 42A-42A strain Y; cn bw sp. *** SEQUENCING IN PROGRESS
***, 97 unordered pieces.
AC008339
AC008339.1 GI:5670396
HTG: HTGS_PHASE1.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 118474)
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shrir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 118474)
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shrir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
```


TITLE	JOURNAL	COMMENT
Rubin, G. M.	Direct Submission	
Submitted (02-AUG-1999)	Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA	
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases.		
* NOTE: This is a 'working draft' sequence. It currently		
* consists of 97 contigs. The true order of the pieces		
* is not known and their order in this sequence record is		
* arbitrary. Gaps between the contigs are represented as		
* runs of N, but the exact sizes of the gaps are unknown.		
* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
* 1	1154: contig of 1154 bp in length	
* 1155	1234: gap of unknown length	
* 1235	1978: contig of 744 bp in length	
* 1979	2058: gap of unknown length	
* 2059	2763: contig of 705 bp in length	
* 2764	2843: gap of unknown length	
* 2844	3474: contig of 631 bp in length	
* 3475	3554: gap of unknown length	
* 3555	4790: contig of 1236 bp in length	
* 4791	4870: gap of unknown length	
* 4871	5691: contig of 821 bp in length	
* 5692	5771: gap of unknown length	
* 5772	7104: contig of 1333 bp in length	
* 7105	7184: gap of unknown length	
* 7185	8019: contig of 835 bp in length	
* 8020	8099: gap of unknown length	
* 8099	8783: contig of 684 bp in length	
* 8784	8863: gap of unknown length	
* 8864	9614: contig of 751 bp in length	
* 9615	9694: gap of unknown length	
* 9695	10352: contig of 658 bp in length	
* 10353	10432: gap of unknown length	
* 10433	11036: contig of 604 bp in length	
* 11037	11116: gap of unknown length	
* 11117	11859: contig of 743 bp in length	
* 11860	11939: gap of unknown length	
* 11940	12630: contig of 691 bp in length	
* 12631	12710: gap of unknown length	
* 12711	13410: contig of 700 bp in length	
* 13411	13490: gap of unknown length	
* 13491	14724: contig of 1234 bp in length	
* 14725	14804: gap of unknown length	
* 14805	15537: contig of 733 bp in length	
* 15538	15617: gap of unknown length	
* 15618	16875: contig of 1258 bp in length	
* 16876	16955: gap of unknown length	
* 16956	18332: contig of 1377 bp in length	
* 18333	18412: gap of unknown length	
* 18413	19139: contig of 727 bp in length	
* 19140	19219: gap of unknown length	
* 19220	20179: contig of 960 bp in length	
* 20180	20259: gap of unknown length	
* 20260	21313: contig of 1054 bp in length	
* 21314	21393: gap of unknown length	
* 21394	22510: contig of 1117 bp in length	
* 22511	22590: gap of unknown length	
* 22591	23469: contig of 879 bp in length	
* 23470	23549: gap of unknown length	
* 23550	24814: contig of 1265 bp in length	
* 24815	24894: gap of unknown length	
* 24895	26082: contig of 1188 bp in length	
* 26083	26162: gap of unknown length	
* 26163	27865: contig of 1703 bp in length	
* 27866	27945: gap of unknown length	
* 27946	29037: contig of 1092 bp in length	
* 29038	29117: gap of unknown length	
* 29118	30403: contig of 1286 bp in length	
* 30404	30483: gap of unknown length	
* 30484	31874: contig of 1391 bp in length	
* 31875	31954: gap of unknown length	
* 31955	33181: contig of 1227 bp in length	
* 33182	33261: gap of unknown length	
* 33262	34653: contig of 1392 bp in length	
* 34654	34733: gap of unknown length	
* 34734	36014: contig of 1281 bp in length	
* 36015	36094: gap of unknown length	
* 36095	37591: contig of 1497 bp in length	
* 37592	37671: gap of unknown length	
* 37672	38920: contig of 1249 bp in length	
* 38921	39000: gap of unknown length	
* 39001	40308: contig of 1308 bp in length	
* 40309	40388: gap of unknown length	
* 40389	41517: contig of 1129 bp in length	
* 41518	41597: gap of unknown length	
* 41598	43130: contig of 1533 bp in length	
* 43131	43210: gap of unknown length	
* 43211	44533: contig of 1323 bp in length	
* 44534	44613: gap of unknown length	
* 44614	45686: contig of 1073 bp in length	
* 45687	45766: gap of unknown length	
* 45767	47157: contig of 1391 bp in length	
* 47158	47237: gap of unknown length	
* 47238	48766: contig of 1529 bp in length	
* 48767	48846: gap of unknown length	
* 48847	50558: contig of 1712 bp in length	
* 50559	50638: gap of unknown length	
* 50639	52587: contig of 1949 bp in length	
* 52588	52667: gap of unknown length	
* 52668	54127: contig of 1460 bp in length	
* 54128	54207: gap of unknown length	
* 54208	56171: contig of 1964 bp in length	
* 56172	56251: gap of unknown length	
* 56252	57717: contig of 1466 bp in length	
* 57718	57797: gap of unknown length	
* 57798	59216: contig of 1419 bp in length	
* 59217	59296: gap of unknown length	
* 59297	61485: contig of 2189 bp in length	
* 61486	61565: gap of unknown length	
* 61566	63346: contig of 1781 bp in length	
* 63347	63426: gap of unknown length	
* 63427	65078: contig of 1652 bp in length	
* 65079	65158: gap of unknown length	
* 65159	67680: contig of 2522 bp in length	
* 67681	67760: gap of unknown length	
* 67761	69367: contig of 1607 bp in length	
* 69368	69447: gap of unknown length	
* 69448	72886: contig of 3439 bp in length	
* 72887	72966: gap of unknown length	
* 72967	74413: contig of 1447 bp in length	
* 74414	74493: gap of unknown length	
* 74494	76339: contig of 1846 bp in length	
* 76340	76419: gap of unknown length	
* 76420	80413: contig of 3994 bp in length	
* 80414	80493: gap of unknown length	
* 80494	88132: contig of 7639 bp in length	
* 88133	88212: gap of unknown length	
* 88213	88883: contig of 671 bp in length	
* 88884	88963: gap of unknown length	
* 88964	89611: contig of 648 bp in length	
* 89612	89691: gap of unknown length	
* 89692	90372: contig of 681 bp in length	
* 90373	90452: gap of unknown length	
* 90453	91133: contig of 681 bp in length	
* 91134	91213: gap of unknown length	
* 91214	91941: contig of 728 bp in length	
* 91942	92021: gap of unknown length	
* 92022	92785: contig of 764 bp in length	
* 92786	92865: gap of unknown length	
* 92866	93548: contig of 683 bp in length	

```

* 93549 93628: gap of unknown length
* 93629 94374: contig of 746 bp in length
* 94375 94454: gap of unknown length
* 94455 95143: contig of 689 bp in length
* 95144 95223: gap of unknown length
* 95224 95850: contig of 627 bp in length
* 95851 95930: gap of unknown length
* 95931 96666: contig of 736 bp in length
* 96667 96746: gap of unknown length
* 96747 97467: contig of 721 bp in length
* 97468 97547: gap of unknown length
* 97548 98374: contig of 827 bp in length
* 98375 98454: gap of unknown length
* 98455 99170: contig of 716 bp in length
* 99171 99250: gap of unknown length
* 99251 99955: contig of 705 bp in length
* 99956 100035: gap of unknown length
* 100036 100860: contig of 825 bp in length
* 100861 100940: gap of unknown length
* 100941 101652: contig of 712 bp in length
* 101653 101732: gap of unknown length

Query Match      3.7%  Score 153.2; DB 41; Length 118474;
Best Local Similarity 98.1%; Pred. No. 2.2e-34;
Matches 155; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtgtcgctgacttttgttaggtgaaggaggcggtggccaaatagtttttggtatacag 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 103315 GTGTGCGTGAAGTGTAGGTGAAGGAGGCGGTGGCCAAATAGTTTGGGTATACG 103374
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 61 gatagaatttggatgataaaacaaacaaacatttttccaaagcgtggaagttt 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 103375 GATAGAAATTTGGATGATAAATAAACCAGCAATCAAAACATTTTCAAAGCGTGAAGTTT 103434
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 121 tggccgcttggggcgaatgcaaaaagtttttgggcta 158
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 103435 TGGCCGCTTGTGGCATGGCAAAACGTTTGTGGGTA 103472
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 11
AC009255/c
LOCUS
DEFINITION
Drosophila melanogaster chromosome 2 clone BACR39E09 (D1043)
RPCI-98 39.E.9 map 42A-42A strain y; cn bw sp. *** SEQUENCING IN
PROGRESS *** 103 unordered pieces.
AC009255
AC009255.3 GI:6554247
HTG: HTGS_PHASE1.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 109699)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 109699)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

```

```

Rubin,G.M.
Direct Submission
Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 10, 1999 this sequence version replaced gi:5734580.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 103 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1199: contig of 1199 bp in length
* 1200 1279: gap of unknown length
* 1280 1997: contig of 718 bp in length
* 1998 2077: gap of unknown length
* 2078 2686: contig of 609 bp in length
* 2687 2766: gap of unknown length
* 2767 3839: contig of 1073 bp in length
* 3840 3919: gap of unknown length
* 3920 4568: contig of 649 bp in length
* 4569 4648: gap of unknown length
* 4649 5320: contig of 672 bp in length
* 5321 5400: gap of unknown length
* 5401 6335: contig of 935 bp in length
* 6336 6416: gap of unknown length
* 6416 7390: contig of 975 bp in length
* 7391 7470: gap of unknown length
* 7471 8350: contig of 880 bp in length
* 8351 8431: gap of unknown length
* 8431 9510: contig of 1080 bp in length
* 9511 9590: gap of unknown length
* 9591 10436: contig of 846 bp in length
* 10437 10516: gap of unknown length
* 10517 11243: contig of 727 bp in length
* 11244 11323: gap of unknown length
* 11324 11946: contig of 623 bp in length
* 11947 12026: gap of unknown length
* 12027 12701: contig of 675 bp in length
* 12702 12781: gap of unknown length
* 12782 13062: contig of 281 bp in length
* 13063 13142: gap of unknown length
* 13143 14057: contig of 915 bp in length
* 14058 14138: gap of unknown length
* 14138 14805: contig of 668 bp in length
* 14806 14885: gap of unknown length
* 14886 15334: contig of 449 bp in length
* 15335 15414: gap of unknown length
* 15415 16203: contig of 789 bp in length
* 16204 16284: gap of unknown length
* 16284 17032: contig of 749 bp in length
* 17033 17112: gap of unknown length
* 17113 18667: contig of 1555 bp in length
* 18668 18747: gap of unknown length
* 18748 19962: contig of 1214 bp in length
* 19962 20041: gap of unknown length
* 20042 20778: contig of 737 bp in length
* 20779 20858: gap of unknown length
* 20859 21803: contig of 945 bp in length
* 21804 22739: contig of 856 bp in length
* 22740 22819: gap of unknown length
* 22820 23590: contig of 771 bp in length
* 23591 23670: gap of unknown length
* 23671 24575: contig of 905 bp in length
* 24576 24655: gap of unknown length
* 24656 25422: contig of 767 bp in length

```

TITLE
JOURNAL
COMMENT

* 25423 25502: gap of unknown length
* 25503 contig of 674 bp in length
* 26177 26256: gap of unknown length
* 26257 26758: contig of 502 bp in length
* 26759 26838: gap of unknown length
* 26839 27932: contig of 1094 bp in length
* 27933 28012: gap of unknown length
* 28013 29308: contig of 1296 bp in length
* 29309 29388: gap of unknown length
* 29389 30496: contig of 1108 bp in length
* 30497 30576: gap of unknown length
* 30577 31728: contig of 1152 bp in length
* 31729 31808: gap of unknown length
* 31809 33080: contig of 1272 bp in length
* 33081 33160: gap of unknown length
* 33161 34333: contig of 1173 bp in length
* 34334 34413: gap of unknown length
* 34414 35478: contig of 1065 bp in length
* 35479 35558: gap of unknown length
* 35559 36465: contig of 907 bp in length
* 36466 36545: gap of unknown length
* 36546 37799: contig of 1154 bp in length
* 37700 37779: gap of unknown length
* 37780 38966: contig of 1187 bp in length
* 38967 39046: gap of unknown length
* 39047 40514: contig of 1468 bp in length
* 40515 40594: gap of unknown length
* 40595 41900: contig of 1306 bp in length
* 41901 41980: gap of unknown length
* 41981 43438: contig of 1458 bp in length
* 43439 43518: gap of unknown length
* 43519 44606: contig of 1088 bp in length
* 44607 44686: gap of unknown length
* 44687 46717: contig of 2031 bp in length
* 46718 46797: gap of unknown length
* 46798 48201: contig of 1404 bp in length
* 48202 48281: gap of unknown length
* 48282 49781: contig of 1500 bp in length
* 49782 49861: gap of unknown length
* 49862 51105: contig of 1244 bp in length
* 51106 51185: gap of unknown length
* 51186 52319: contig of 1134 bp in length
* 52320 52399: gap of unknown length
* 52400 54918: contig of 2519 bp in length
* 54919 54998: gap of unknown length
* 54999 56678: contig of 1680 bp in length
* 56679 56758: gap of unknown length
* 56759 59360: contig of 2602 bp in length
* 59361 59440: gap of unknown length
* 59441 62856: contig of 3416 bp in length
* 62857 62936: gap of unknown length
* 62937 66768: contig of 3832 bp in length
* 66769 66848: gap of unknown length
* 66849 70379: contig of 3531 bp in length
* 70380 70459: gap of unknown length
* 70460 78725: contig of 8266 bp in length
* 78726 78805: gap of unknown length
* 78806 79327: contig of 522 bp in length
* 79328 79407: gap of unknown length
* 79408 80062: contig of 655 bp in length
* 80063 80142: gap of unknown length
* 80143 80637: contig of 495 bp in length
* 80638 80717: gap of unknown length
* 80718 81225: contig of 508 bp in length
* 81226 81305: gap of unknown length
* 81306 81994: contig of 689 bp in length
* 81995 82074: gap of unknown length
* 82075 82836: contig of 762 bp in length
* 82837 82916: gap of unknown length
* 82917 83557: contig of 641 bp in length
* 83558 83637: gap of unknown length
* 83638 83920: contig of 283 bp in length
* 83921 84000: gap of unknown length

* 84001 84514: contig of 514 bp in length
* 84515 84594: gap of unknown length
* 84595 85404: contig of 810 bp in length
* 85405 85484: gap of unknown length
* 85485 86089: contig of 605 bp in length
* 86090 86169: gap of unknown length
* 86170 86848: contig of 679 bp in length
* 86849 86928: gap of unknown length
* 86929 87211: contig of 283 bp in length
* 87212 87291: gap of unknown length
* 87292 87990: contig of 699 bp in length
* 87991 88070: gap of unknown length
* 88071 88629: contig of 559 bp in length
* 88630 88709: gap of unknown length
* 88710 89373: contig of 664 bp in length
* 89374 89453: gap of unknown length
* 89454 89767: contig of 314 bp in length
* 89768 89847: gap of unknown length
* 89848 90407: contig of 560 bp in length
* 90408 90487: gap of unknown length
* 90488 90961: contig of 474 bp in length
* 90962 91041: gap of unknown length

Query Match 3.3%; Score 137; DB 42; Length 109699;
Best Local Similarity 58.7%; Pred. No. 1.9e-29;
Matches 215; Conservative 0; Mismatches 86; Indels 65; Gaps 1;

QY 415 caaacagtggttcccgaggagcgagggaatgcagcctatgatgcgttctgcgcgcg 474
DB 83918 CGAACAGCTGTTCTCCCGTGGACGGAGGAAATTCAGCTTATGGATGGTTTCGTGCGCGC 83859

QY 475 ctacacgaagtttttaattgctgaattgtcgcaatcaccagcagcagcaaaa---- 530
DB 83858 CTACACGAAGTTTTTAAATTGCTTGAAATTGTCGCAATCACCACCAACCAAGTGA 83799

QY 531 ----- 530

DB 83798 GTACATCCAAAGGCTCTCCAGTCTCTCAACAACACTGATTACTTAACTATGCGACA 83739

QY 531 -caaacgtacctggagcacaaactgcctacatgtagacaacgtaaagcgcggaacacag 589

DB 83738 GCAACGTCACCCCTGGAGCACAACTGCTACATGTAGACAACGTAAAGCGCGGAACACAG 83679

QY 590 ggcagctggaccagatcttccggagtgatgcgtgcgtgcgtgcgcgcgcacacctgt 649
DB 83678 GGCAGCTGGACCAGATCTTTCCGGAGTATGGCTGCCCTGCTGNNNNNNNNNNNNNN 83619

QY 650 ggacgcagaactctcagaactttactcgggacacaaacatcctgaacacgatatattcagt 709

DB 83618 NNN 83559

QY 710 accata 715
DB 83558 NTAATAA 83553

RESULT 12
AC017232/c
LOCUS AC017232 10844 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017232
VERSION AC017232.1 GI:6553754
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 10844)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,

COMMENT	Rockville, MD, USA	
	This sequence was identified as CDM:10209962 by the submitter.	
	For more information on this record e-mail to fly@celera.com.	
FEATURES	* NOTE: This is a 'working draft' sequence.	
	* This sequence will be replaced	
	* by the finished sequence as soon as it is available and	
source	* the accession number will be preserved.	
	Location/Qualifiers	
	l..10844	
BASE COUNT	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	3202 a 2155 c 2105 g 3382 t	
ORIGIN	Query Match 1.5%; Score 62.4; DB 42; Length 10844;	
	Best Local Similarity 75.3%; Pred. NO. 6e-07;	
	Matches 110; Conservative 0; Mismatches 21; Indels 15; Gaps 2;	
QY	8 tgactgtttgttagtgtaaggagggcggtggccaaatagtttttgatcacgtagaa 67	
	Db 5920 TGGGTGTTTGTAGGTAGTAGGAGGGCGGTGGCAAAAGTGTTTTGTATACCGACAGAA 5861	
QY	68 ttggatga-----aaataaaacgaatacaaacattttttcaaaa-----gcgt 112	
	Db 5860 ATTGGTAAACAAACAATAAAACCAACAAAATCAAAACATTTTCAAAAGTGTGGCGGT 5801	
QY	113 ggaagtgttggcggtgttgccat 138	
	Db 5800 AGCAGTTTTGGGAGGCTGTGGCGGT 5775	
RESULT 13	AE003066 28380 bp DNA INV 22-MAR-2000	
	Drosophila melanogaster genomic scaffold 14200001385325, complete	
	sequence.	
ACCESSION	AE003066	
	AE003066.1 GI:7289027	
	HTG.	
KEYWORDS	fruit fly.	
	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
REFERENCE	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
	1 (bases 1 to 28380)	
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,	
	Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,	
	George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,	
	Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,	
	Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,	
	Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor	
	Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,	
	Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,	
	Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,	
	Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,	
	Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,	
	Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,	
	Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de	
	Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,	
	Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,	
	Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,	
	Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,	
	Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,	
	Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,	
	Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,	
	Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,	
	Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,	
	Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,	
	Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,	
	McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,	
	Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,	
	Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,	

TITLE	Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,	
	Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,	
	Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klamos,I.,	
JOURNAL	Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,	
	Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,	
	Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,	
MEDLINE	Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,	
	Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,	
	Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,	
AUTHORS	Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,	
	Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.	
	The genome sequence of Drosophila melanogaster	
TITLE	Science 287 (5461), 2185-2195 (2000)	
	20196006	
	2 (bases 1 to 28380)	
JOURNAL	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
	Direct Submission	
	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,	
MEDLINE	Rockville, MD, USA	
	Location/Qualifiers	
	1..28380	
FEATURES	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	complement(join(5607..6323,6668..6957,7095..7257))	
source	/gene="CG13865"	
	/product="CT33380"	
	/db_xref="FLYBASE:FBan0013865"	
mRNA	/db_xref="FLYBASE:FBgn0039965"	
	complement(<5607..>7257)	
	/gene="CG13865"	
gene	/db_xref="FLYBASE:FBan0013865"	
	/db_xref="FLYBASE:FBgn0039965"	
	complement(join(5913..6323,6668..6957,7095..7194))	
CDS	/gene="CG13865"	
	/note="CG13865 gene product"	
	/codon_start=1	
	/db_xref="FLYBASE:FBan0013865"	
	/db_xref="FLYBASE:FBgn0039965"	
	/protein_id="AAF45339.1"	
mRNA	/db_xref="GI:7289028"	
	/translation="MAYDKKESDVKEVYKLVGRFVCYSEKKPEACHLLGDYLEG	
	IKKDFKASKYKSTCDDYKASKYKYNISFLGKSGKSGKGNPOVAYEYKGNL	
	NDSADCLHSLGLLVKSMPEIDWNVPKLEFLTKSDLNNAACFYLSGMHISGVOK	
	KQDSAVTASSGSGTSPGAPQPKLDSYIVLKKMKKAFQFAHKACELRNMYACANL	
	SOMYARGDGIETKEKEAEYKKLALQEMDEVKKQHDTLGFGQGVGMFN"	
	join(7547..7733,7795..>8399)	
	/gene="CG17515"	
	/product="CT33382"	
gene	/db_xref="FLYBASE:FBan0017515"	
	/db_xref="FLYBASE:FBgn0039966"	
	<7547..>8399	
CDS	/gene="CG17515"	
	/note="CG17515 gene product"	
	/codon_start=1	
	/db_xref="FLYBASE:FBan0017515"	
	/db_xref="FLYBASE:FBgn0039966"	
	join(7670..7733,7795..8399)	
	/gene="CG17515"	
	/note="CG17515 gene product"	
	/translation="MSSSRTRNGPTLNFRAVLLEGCVGKTSLVLYRMEDRPNQAHLS	
	TLQASPVSRKMSLEDRRAQINWDTAGOEPRHALGPIYVRGSDGALLVYDITDRDSF	
	OKVKSVMRELQMRGTEIALIVGNKTDLEORAVTHDEALQVARTVGQVYVTSKAE	
	NEGVAFELFELLTLMLEQLSQRPDASPLRLQNPDTDLNNSDDSEAPDGPAGORS	
mRNA	CCG1"	
	join(<24856..25554,25613..26131,26612..27888)	
	/gene="stnA"	
	/note="Nucleotide sequence of the Celera sequence differs	
	from the published sequence for this transcript."	
	/product="CT33379"	
	/db_xref="FLYBASE:FBan0012500"	

```
/db_xref=FLYBASE:FBgn0016976"
/evidence=not_experimental
<24856..>27888
/gene="stnA"
/note="CG12500"
/map=20A4-20A5"
/db_xref=FLYBASE:FBan0012500"
/db_xref=FLYBASE:FBgn0016976"
/evidence=not_experimental
join(24856..25554,25613..25663)
/gene="stnA"
/note="stnA gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref=FLYBASE:FBan0012500"
/db_xref=FLYBASE:FBgn0016976"
/evidence=not_experimental
/protein_id="AAF45340.1"
/db_xref="GI:7289029"
/translation="MLKLPGKLKKKKKKDOELTFEELEQYKRDLLKAKOEAAATK
SDAGESDGSDDVAHPEVAFNSCLGSGSSSILNAQOOLSDQNOGAGDEKAF
KALTSQVDSILHKQTDELDRINKESFYORLPSAAEKKKQKEEAARLAEQDEKQR
LQQLAEQRDKLAEAVOLSESEEEAGDYADDIFATDYIEAITSGLQLAVVPDSPVL
AEDGPDPDFDTAYAEKPLRFYPVAVTESTLSP"
BASE COUNT      6404 a 4440 c 4208 g 6616 t 6712 others
ORIGIN

Query Match      1.5%; Score 62.4; DB 34; Length 28380;
Best Local Similarity 75.3%; Pred. No. 7.5e-07;
Matches 110; Conservative 0; Mismatches 21; Indels 15; Gaps 2;

Qy      8  tqagcttttgtagtgtaagaggagggcgtagccaaatagtttttgggtatagcagatagaa 67
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      9428  TCGGCGCTTTTGTAGGTGTAAGGAGGGCGTGGCAAAAGTGTGTTTGGTATACCGACAGAA 9369

Qy      68  ttggtatga-----aaataacgaacgaatacaaacatttttccaaa-----ggcgtgggaagt 112
      ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db      9368  ATTGTTAAACAAACAATAAACCAGAAATAATCAAAACATTTTTCAAAAGTGTGGCGT 9309

Qy      113  ggaagtttggccgcttgtagggcat 138
      ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db      9308  ACCAGCTTTGGGAGGCTTGCTGGCGCT 9283

RESULT 14
AC017996
LOCUS
DEFINITION      Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC017996
AC017996.1 GI:6553194
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 16047)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212940 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..16047
/organism="Drosophila melanogaster"
```

```
/db_xref="Taxon:7227"
BASE COUNT      4912 a 3114 c 3241 g 4780 t
ORIGIN

Query Match      1.5%; Score 61.4; DB 42; Length 16047;
Best Local Similarity 75.5%; Pred. No. 1.3e-06;
Matches 105; Conservative 0; Mismatches 26; Indels 8; Gaps 2;

Qy      8  tqagcttttgtagtgtaagaggagggcgtagccaaatagtttttgggtatagcagatagaa 67
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      6827  TCGGCGCTTTTGTAGGTGTAAGGAGGGCGTGGCCAAATGTTTGTAGTATACCGATTGAA 6886

Qy      68  ttggtatga--aaataacgaacgaatacaaacatttttccaaa-----aggcgtgggaagt 119
      ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db      6887  ATTCAAACAATAATACTGAGAAAAATCAAAACACTTTTCAAAAGTGTGACCGTGGCAGTT 6946

Qy      120  ttggcgcgcttgtagggcat 138
      ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db      6947  TTGGTCTGTGTTCTGGACGT 6965

RESULT 15
AC003781
LOCUS
DEFINITION      Drosophila melanogaster genomic scaffold 14200001386030 section 1
of 3, complete sequence.
AC003781 AE002725
VERSION      HTG.
KEYWORDS      fruit fly.
SOURCE      Drosophila melanogaster
ORGANISM      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 311786)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
Sutton,C.G., Wortman,J.R., Richards,S., Ashburner,M., Henderson,S.N.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Zhang,Q., Chen,L.X.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Champen,M., Pfeiffer,B.D.,
Miklos,G.L., Abrial,J.F., Aqbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Folsler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Murny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
```



```
/product="CT11930"  
/db_xref="FLYBASE:FBan0003549"  
/db_xref="FLYBASE:FBgn0032954"  
<36269> >37921  
/gene="CG3549"
```

gene

```
Query Match      1.5%; Score 61.4; DB 58; Length 311786;  
Best Local Similarity 75.5%; Pred. No. 2.7e-06;  
Matches 105; Conservative 0; Mismatches 26; Indels 8; Gaps 2;  
  
QY      8  tgactgtttttagtgtaagaggggcgtagccaaatagtttttgggtatacggatagaa 67  
      ||  ||||||||||||||||||||||||||||||||||||||||||||||||  
Db 163121 TGGGCGTTTTGTAGGTGTAAGGAGGGCGTGGCCAAATTGTTTAGTATACCGATTGAA 163180  
  
QY      68  ttggatga--aaaataaaacgaatacaaacatttttcaaa-----agcgtgaaagtt 119  
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  
Db 163181 ATTCAACAATATACTGAGAAAAATCAAAACACTTTTCAAAGTGTGAGCGTGGCAGTT 163240  
  
QY      120 ttggcgcgcttgtgggcat 138  
      |||  ||  ||  ||  ||  ||  
Db 163241 TTGGTCTGTTTGTGGACGT 163259
```

Search completed: September 4, 2000, 00:48:39
Job time: 22339 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:24:48 ; Search time 4457.97 Seconds
(without alignments)
4083.422 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgtgccgactgttttgtta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 5247842 seqs, 2204914090 residues 10495684
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	EST:*
	1: em_est1:*
	2: em_est2:*
	3: em_est3:*
	4: em_est4:*
	5: em_est5:*
	6: em_est6:*
	7: em_est7:*
	8: em_est8:*
	9: em_est9:*
	10: em_est10:*
	11: em_est11:*
	12: em_est12:*
	13: em_est13:*
	14: em_est14:*
	15: em_est15:*
	16: em_est16:*
	17: em_est17:*
	18: em_est18:*
	19: em_est19:*
	20: gb_est1:*
	21: gb_est2:*
	22: gb_est3:*
	23: gb_est4:*
	24: gb_est5:*
	25: gb_est6:*
	26: gb_est7:*
	27: gb_est8:*
	28: gb_est9:*
	29: gb_est10:*
	30: gb_est11:*
	31: gb_est12:*
	32: gb_est13:*
	33: gb_est14:*
	34: gb_est15:*
	35: gb_est16:*
	36: gb_est17:*
	37: gb_est18:*
	38: gb_est19:*
	39: gb_est20:*
	40: gb_est21:*
	41: gb_est22:*
	42: gb_est23:*
	43: gb_est24:*
	44: gb_est25:*
	45: gb_est26:*
	46: gb_est27:*
	47: gb_est28:*
	48: gb_est29:*
	49: gb_est30:*
	50: gb_est31:*
	51: gb_est32:*
	52: em_est20:*
	53: em_est21:*
	54: em_est22:*
	55: em_est23:*
	56: em_est24:*
	57: em_est25:*
	58: em_est26:*
	59: gb_est33:*
	60: gb_est34:*
	61: gb_est35:*
	62: gb_est36:*
	63: gb_est37:*
	64: gb_est38:*
	65: em_est27:*
	66: em_est28:*
	67: em_est29:*
	68: em_est30:*
	69: gb_est39:*
	70: gb_est40:*
	71: gb_est41:*
	72: gb_est42:*
	73: gb_est43:*
	74: gb_est44:*
	75: em_est31:*
	76: em_est32:*
	77: em_est33:*
	78: em_est34:*
	79: gb_est45:*
	80: gb_est46:*
	81: gb_est47:*
	82: em_est35:*
	83: em_est36:*
	84: em_est37:*
	85: gb_est48:*
	86: gb_est49:*
	87: gb_est50:*
	88: gb_est51:*
	89: gb_est52:*
	90: gb_est53:*
	91: gb_est54:*
	92: gb_est55:*
	93: gb_gss1:*
	94: gb_gss2:*
	95: gb_gss3:*
	96: gb_gss4:*
	97: em_gss1:*
	98: em_gss2:*
	99: em_gss3:*
	100: em_gss4:*
	101: gb_gss5:*
	102: gb_gss6:*
	103: gb_gss7:*
	104: gb_gss8:*
	105: gb_gss9:*
	106: em_gss5:*
	107: em_gss6:*
	108: em_gss7:*
	109: em_gss8:*
	110: em_gss9:*
	111: em_gss10:*
	112: em_gss11:*
	113: gb_gss10:*
	114: gb_gss11:*
	115: em_gss12:*
	116: gb_gss12:*

117: gb_gss13:.*
 118: gb_gss14:.*
 119: gb_gss15:.*
 120: gb_gss16:.*
 121: gb_gss17:.*
 122: gb_gss18:.*
 123: gb_gss19:.*
 124: em_gss13:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	363.4	8.8	754	CNS0116H	AL099875 Drosophila
c 2	321.6	7.8	328	46	A1945842
c 3	99.6	2.4	390	69	AW240273
c 4	80.4	1.9	546	79	AW654475
c 5	73.2	1.8	399	88	R55122
c 6	61.2	1.5	561	27	AA498678
c 7	61.2	1.5	432	21	AA106247
c 8	60.4	1.5	647	23	AA254928
c 9	60.2	1.5	611	47	AL043581
c 10	51.8	1.3	333	25	AA354156
c 11	50	1.2	1101	122	CNS001U2
c 12	49.4	1.2	1201	123	CNS0163J
c 13	48.4	1.2	1101	122	CNS00AX2
c 14	48	1.2	1101	122	CNS00ERH
c 15	47.2	1.1	982	122	CNS00EO8
c 16	47	1.1	529	79	AW632482
c 17	46.4	1.1	424	79	AW662299
c 18	46.4	1.1	792	47	AL040564
c 19	46.4	1.1	1101	122	CNS00BMG
c 20	46.2	1.1	360	81	D67504
c 21	46.2	1.1	440	22	AA195538
c 22	45.4	1.1	517	79	AW656527
c 23	45.2	1.1	940	122	CNS00CF0
c 24	45	1.1	959	122	CNS00Z2S
c 25	44.6	1.1	503	81	C89382
c 26	44.4	1.1	312	72	AW417103
c 27	44.4	1.1	1028	122	CNS00LD8
c 28	44.2	1.1	986	122	CNS00KT1
c 29	43.8	1.1	255	80	C18401
c 30	43.8	1.1	326	25	AA378263
c 31	43.8	1.1	448	47	AL045225
c 32	43.8	1.1	511	22	AA195161
c 33	43.8	1.1	634	47	AL043582
c 34	43.8	1.1	832	122	CNS007XU
c 35	43.4	1.1	561	36	AL218269
c 36	43.4	1.1	623	72	AW439423
c 37	43.2	1.0	429	64	AW125000
c 38	43.2	1.0	446	22	AA217968
c 39	43.2	1.0	449	20	AA049660
c 40	43.2	1.0	491	29	AA625370
c 41	43.2	1.0	491	36	AA154968
c 42	43.2	1.0	492	26	AA422332
c 43	43.2	1.0	515	32	AA870601
c 44	43.2	1.0	1101	122	CNS00LHX
c 45	43	1.0	465	35	AA145323

ALIGNMENTS

RESULT 1
 CNS0116H/c
 LOCUS CNS0116H 754 bp DNA GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN05M24 of DrosBAC library from Drosophila melanogaster (fruit

ACCESSION
 AL099875
 VERSION
 AL099875.1
 KEYWORDS
 GSS.
 SOURCE
 fruit fly.
 ORGANISM
 Drosophila melanogaster

fly), genomic survey sequence.

AL099875.1 GI:5611486

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

1 (bases 1 to 754)

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelOBAC11.

Location/Qualifiers

1. 754

/organism="Drosophila melanogaster"

/plasmid="pBelOBAC11"

/db_xref="taxon:7227"

/clone_lib="DrosBAC"

/clone="BACN05M24"

/note="end : T7"

BASE COUNT 198 a 129 c 183 g 217 t 27 others

ORIGIN

Query Match 8.8%; Score 363.4; DB 122; Length 754;

Best Local Similarity 94.2%; Pred. No. 4.8e-98;

Matches 359; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 150 ttttgctatccgttaatacaataacggttgccgggacaataaccaccacgaagctgtgt 209

Db 438 TTTTCACTATCCTTTAATCAAAATACCGTTGCCCGGAAATACCCACCAAGATCGGTGT 379

QY 210 acctacgaactggatccggtatcgctgtcatggtctcttaatacatcctcgactac 269

Db 378 ACCTACRAAAGTGGATCCGGATCGTGTGTCATGCGACTCTTTTAATACATCCTCGACTAC 319

QY 270 accgcaggaaacgcacccctcccggaacacctggcccccccgaaaccacaggtactcaatag 329

Db 318 ACCGCAGGAACCGCMCCCTTCCGGCGAAACCGTCCGCCCGCCVMAACACAGTACTMAATAG 259

QY 330 cagtacacgacccgacgacctccctcttgcctggcgacagagccaccctt 389

Db 258 CAGTACACGAGACCGGACGCGCGCTCCCTTCCTGCGCTGGGCKTWTMGACACCCCGCYTT 199

QY 390 ttctacgtccacagattactctggaacacagttctccctggacgaggaatgca 449

Db 198 TTTTCTAGTCCACAGATTACTCTGCGAACACAGTGTCTCCCGTGGACGAGGGAATGCA 139

QY 450 gctatgtagtgcgttcgtgctgctgacgaagtttttaattgctgaaattgtgog 509

Db 138 GCTTATGATGCGTTTCGTGCGCGCTACAGAAKTTTTTAAATTGCTTGAATTTGTCG 79

QY 510 caatcaccagacgacgaaaa 530

Db 78 CAATACACGAGCAGCGAAAA 58

RESULT 2
 AI945842
 LOCUS AI945842 328 bp mRNA EST 17-AUG-1999
 DEFINITION bs17e03.y1 Drosophila melanogaster adult testis library Drosophila melanogaster cDNA clone bs17e03 5', mRNA sequence.

```

ACCESSION      AI945842
VERSION        AI945842.1  GI:5736240
KEYWORDS      EST.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 328)
AUTHORS        Andrews, J., Bouffard, G. and Oliver, B.
TITLE          Drosophila melanogaster testis expressed sequence tags
JOURNAL        Unpublished (1999)
COMMENT        Laboratory of Cellular and Developmental Biology
               NIDDK, National Institutes of Health
               6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
               Fax: (301) 496 5239
               Email: oliver@helix.nih.gov,
               http://www.nidk.nih.gov/intram/people/boliver.htm
               Tissue isolation and library construction performed at the National
               Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
               http://www.nidk.nih.gov/intram/people/boliver.htm). DNA sequencing
               and analyses performed by National Institutes of Health Intramural
               Sequencing Center (NISC; see http://www.nisc.nih.gov). Chromatogram
               data were analyzed and evaluated for high quality using the ted
               program (Gleeson T and Hillier L, 1991).
               Plate: 17 row: e column: 03
               Seq primer: M13R1 reverse primer (ABI).
FEATURES       source
               Location/Qualifiers
               1..328
               /organism="Drosophila melanogaster"
               /strain="y[*] w[67cl]/Y"
               /db_xref="taxon:7227"
               /clone="bs17e03"
               /clone_lib="Drosophila melanogaster adult testis library"
               /sex="male"
               /dev_stage="1-5 day adult"
               /lab_host="SOLR (Stratagene)"
               /note="Organ: testis; Vector: pBluescript SK (Stratagene);
               Site_1: EcoRI; Site_2: Xho I; Testes dissected from 1-5
               day adult y[*] w[67cl]/Y males raised at 250C. RNA
               isolated using Trizol (Life Technologies) and a single
               round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
               library constructed using Stratagene ZAP-cDNA synthesis
               kit. Oligo dt-primered, size fractionated -1-6 kb, and
               directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
               Following a single round of amplification pBluescript SK
               phagemids were mass excised. A distribution channel for
               clones is being sought, but not currently available.
               Requests for clones cannot be honored."
               95 a 77 c 91 g 65 t
BASE COUNT     95 a 77 c 91 g 65 t
ORIGIN
Query Match    7.8%; Score 321.6; DB 46; Length 328;
Best Local Similarity 98.8%; Pred. No. 1.4e-85;
Matches 324; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2765 gcgcacagcgccgtagaattgtgggaatccctgcgggattgttaagtgtgcatacc 2824
Db 1 GCACGAGCGCGCGCTAGAAATGTGGGAATCCCTGCCGGATTGCTTAAGTGTGCATACC 60

Qy 2825 aggaagacgcgaagaaacacagggtataacccacatccactgaacgcgcgtacgaqta 2884
Db 61 AGGAAGACGCGAAGAAACACAGGGTATAACCCACATCCACCTGAACGCGCATCGAGTGA 120

Qy 2885 ttgtggcgcgtcttaatggccactagatttttaccgccttagacgcgtactacaaggga 2944
Db 121 TTGTCGGCGGCTCTTAATGCCGCCACTAGATTTTACCGCTTAGAGACGACTACAAAGGGA 180

Qy 2945 agcaaatcgactgggggttttacctcggtctacaggagaactcatgttcgaactggtacca 3004
Db 181 ACCAATCGACTGGGGTTTACCTCGGCTTACAGGAGAACTCATGTTCGAACGTGATCCA 240

```

```

Qy 3005 ctggaagcctgggatttaatttcacagcagcgctgtcagcaagaagcatccagaaga 3064
Db 241 CTGGAAGCCTGGGATTAAATGTTTCAGCAGCAGCGCTGTCTCAGCAAGAAGCATCCAGAAGA 300

Qy 3065 ccaccaaggaagaaatgaaaaatcacatt 3092
Db 301 CCACCAAGGAGGAATGAAAATCACATT 328

RESULT 3
LOCUS    AW240273
DEFINITION up29g10.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:2655810 5' similar to TR:P97260 P97260 SRFBP CLEAVAGE ACTIVATING PROTEIN. ;, mRNA sequence.
ACCESSION AW240273
VERSION    AW240273.1  GI:6574096
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 390)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
            Unpublished (1997)
JOURNAL    On Mar 8, 1999 this sequence version replaced gi:4388424.
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            MGI:l033958
            Seq primer: -40RP from Gibco.
FEATURES     Location/Qualifiers
            1..390
            /organism="Mus musculus"
            /strain="FVB-3"
            /db_xref="taxon:10090"
            /clone="IMAGE:2655810"
            /clone_lib="NCI_CGAP_Mam2"
            /tissue_type="tumor, biopsy sample"
            /dev_stage="5 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"
BASE COUNT   79 a 116 c 105 g 88 t 2 others
ORIGIN
Query Match    2.4%; Score 99.6; DB 69; Length 390;
Best Local Similarity 58.4%; Pred. No. 1.2e-18;
Matches 174; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 1208 tggaaaaatagcttggtgatcacaagaagcgctgctccaatggcagacattcgacgtga 1267
Db 2 TAGAGAACCTGTTGGTGCTCACCAGTCAGTGGTATCACTCCAGTCGACCTCGAGGTGA 61

Qy 1268 agatccgcgtggcgcaggctcttagcaagagggttggcatatatatccaagaactctttga 1327
Db 62 AGCTTCGGATTGCACAAGGCTTGAGTAGTGAGAGCTGGTCCATCATGAAGACGCGCGGA 121

Qy 1328 cggagataacaatttttgacaattggtcttgctacttctgtgccgcgtatccaggagtttt 1387

```

```

Db 122 CCGAGCTGGGCATCATCTCATTTGGCTACTTCACCCCTCGCTGCTATCAGGAGTTCT 181
QY 1388 gtatctttgcatagtcggttcttccgattttatgtacagatgctctcttctcaa 1447
Db 182 GCCTCTTTGTTGTGTGGGGCTGTGTCTACTTCTTCCCTCCAGATGCTGTTCTTACCA 241
QY 1448 caatactggccatgaacattaaagcagaccagatatacgcgagggccaagcaccttcc 1505
Db 242 CTGTCTCTGCATCGACATTCGGCGGATGGAGTAGCAGACCTAAACAAGCGGCTGCC 299

RESULT 4
AW654475 546 bp mRNA EST 05-APR-2000
LOCUS 103951 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW654475
VERSION AW654475.1 GI:7420301
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 546)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT On May 6, 1998 this sequence version replaced gi:3114943.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCCGAGTCACGACG
Plate: 88 row: F column: 6
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..546
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 99 a 167 c 165 g 95 t
ORIGIN
Query Match 1.9%; Score 80.4; DB 79; Length 546;
Best Local Similarity 53.3%; Pred. No. 8.1e-13;
Matches 196; Conservative 0; Mismatches 166; Indels 6; Gaps 1;
QY 3109 gccatcagcagcaatcaatcagtcagtcaggtgttaacagacatggtttcactgcagc 3168
Db 32 GCACACAGAAGCCCATCACAGCCCTGAAGCGGCTGCCGGGCGCTCTGTCGACTGGCAGC 91
QY 3169 cagatcacacctcaaggtgtattgctcaataaagtccgattgtgagtacgctccac 3228
Db 92 CAGGACCACACATGACAGTGTTCGGGCTGGAGGATTCTGCTGCGCTCTTCACCCCTGCAG 151
QY 3229 ggtcactgtggcgtgaactgtctctttgttgatcgctggcaacctggcacagggggg 3288

```

```

Db 152 GGCCACTCGGGGCCATCAGGACCGTGTATATTGA-----CCAGACCATGGTGTCTGCC 205
QY 3289 tctgggtccagagcagcgcctctctcgatgtggatctgttcacggagcctcatgtat 3348
Db 206 AGGGGAGGCGCAAGACGGGGCCATCTGCCTGTGGAGCTGTGACCGGCGGTGAGC 265
QY 3349 aatataaagctcaagcagcagcgcgtcagctgcctggtgcctgtgcgccttgcagtaac 3408
Db 266 CACATGTTCCCTCACCGTGGGATGTCACCTCCCTCACCTGCACCACTGTGTGGTCA 325
QY 3409 tcgctagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3468
Db 326 AGCAGTGGCTGGAGCAGCCTCATCAGCATCTGGGACCGCAGCAGCGGCATCAAGCTCTAC 385
QY 3469 accatacaa 3476
Db 386 TCCATCCA 393

RESULT 5
R55122 399 bp mRNA EST 22-MAY-1995
LOCUS y987f07.r1 Soares infant brain lNIB Homo sapiens cDNA clone
DEFINITION IMAGE:40365 5' similar to SP:D2013.8 CE00933 ; , mRNA sequence.
ACCESSION R55122
VERSION R55122.1 GI:824458
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 399)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:803832.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 3115
High quality sequence stops: 302 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3115 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 302.
FEATURES
source
Location/Qualifiers
1..399
/organism="Homo sapiens"
/db_xref="GBB:412906"
/db_xref="taxon:9606"
/clone_lib="IMAGE:40365"
/clone_lib="Soares infant brain lNIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dt) primer [5'
AACTGGAGATTCGCGCGGACGGAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaído."

```

```
BASE COUNT      72 a   120 c   105 g   90 t   12 others
ORIGIN

Query Match      1.8%;   Score 73.2;  DB 88;   Length 399;
Best Local Similarity 57.1%; Pred. No. 1.le-10;
Matches 129; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1280 cgcaggtcttagaagaggttggtgcataatccaaactcttttgacggagataacaa 1339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7  CCCAAGGCCTAAGCAGCAGACTGNTNCCATCATGAAGAAACATGCCACGGAGCTGGGCA 66
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1340 ttttgacaattggtcttctactcttcgtgccgtcatccagaggtttgtgtacttttgcca 1399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 TCATCTCATCGGCTACTTCCACCTAGTGGCGGCCATCCAGAGTCTCTCTNTCTG 126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1400 tagtcggttcttccgattttatgtacagatgtctcttcttcaacaatactggcca 1459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 TCGTGGGCTGTGTCTGACTTCTTCCCTCAGATGCTGTTTTCCACCACCTGTCTGTCCA 186
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1460 tgaacattaaaggaccagatatacggcgaggagcccaacacttcc 1505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 TTGACATTGCCGGATGGAGTAGCAGACCTGAACAAGCGACTNCC 232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AA498678      561 bp   mRNA      EST      01-JUL-1997
LOCUS      vhs3d06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
DEFINITION      clone IMAGE:888779 5' similar to TR:G1228047 G1228047 KIAA0199
PROTEIN      ; mRNA sequence.
ACCESSION      AA498678
VERSION      AA498678.1 GI:2233701
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 561)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:516739
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 377.
Location/Qualifiers
1. .561
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:888779"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue.type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTACGAATCTGAAGTGGAGCGCCCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
[GTGGATTCGTTACCT], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."
```

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 355.

FEATURES

source
1. .432
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:521910"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'

BASE COUNT 97 a 121 c 117 g 97 t
ORIGIN

Query Match 1.5%; Score 61.2; DB 21; Length 432;
Best Local Similarity 50.0%; Pred. No. 4.5e-07;
Matches 184; Conservative 0; Mismatches 178; Indels 6; Gaps 1;

QY 3106 ctagcccatcagcagcaatcacatgcagtcggttaacgacatggttttcaactggc 3165
Db 13 CTGTGAACCAAGAGCCATCACAGCCCTGAGAGCTGCTCCGGGCGCCTAGTGACAGG 72

QY 3166 agccagatcacacctcaagggtgattgctcctaataagtcgagatgttgagatacgtc 3225
Db 73 AGCCAAGACCATACTTAAGAGTCTCCGACTGGATGACTCGTGTGTCCTTTTACCCCTG 132

QY 3226 cagggctcactgtggcctgtaacctgtctcttctgttgatcgctggcaacctggcacagg 3285
Db 133 AAGGGCCACTCAGGGGCAATCACAGCTGTGTACATTGAT-----CAGACCATGGTACTG 186

QY 3286 gggctcgggtccagcagcgccctgctcgtatgagatgctgttcacggagcgtcatg 3345
Db 187 CCAGTGGAGGACAAATGAGGCATCTGCTGTGGATGTACTACAGCAGCCGGGTC 246

QY 3346 tataatacaagctcacgagcgctcagctgctggcctgtgtgcgccagttacgta 3405
Db 247 AGCCAACAATTTGCTACCGTGGAGATGTTACCTCCCTCACCTGTACCGTCTCTGTGTC 306

QY 3406 atctcgtacgacgacgagagatttcggtatggaagcatttcagggaacctgtg 3465
Db 307 ATTAGTAGTGGCTGGATGACTTCATCAGTATCTGGACCGCAGCAGCATCAAGCTG 366

QY 3466 actaccat 3473
Db 367 TACTCCAT 374

RESULT 8

AA254928 647 bp mRNA EST 14-MAR-1997
LOCUS mz84c05.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720104 5' similar to TR:G1228047 G1228047 KIAA0199 PROTEIN ; mRNA sequence.
DEFINITION
AA254928
ACCESSION
AA254928.1 GI:1889532
VERSION
AA254928.1
KEYWORDS
EST.
SOURCE
house mouse.
house mouse.
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 647)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and Waterston, R.

TITLE

The WashU-HHMI Mouse EST Project

JOURNAL COMMENT

Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:445600

Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 438.

FEATURES

source

1. .647
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:720104"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TCTTACCAATCTGAAGTGGAGCGCGCGAATCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 137 a 187 c 177 g 145 t 1 others
ORIGIN

Query Match 1.5%; Score 60.4; DB 23; Length 647;
Best Local Similarity 49.9%; Pred. No. 9.2e-07;
Matches 182; Conservative 0; Mismatches 177; Indels 6; Gaps 1;

QY 3109 gcccatcagcagcgaatcacatgcagtcggttaacgacatggttttcaactggcagc 3168
Db 279 GCACACCAAGAGCCATCACAGCCCTGAGAGCTGCTCCGGGCGCCTAGTGACAGGAGC 338

QY 3169 caggatcacacctcaagggtgattgcctcaataagtcgagatgttgagatacgtccac 3228
Db 339 CAAGACCATACTTAAGAGTCTTCCAGCTGGATGCTCGTGTGCTCTTTACCTCAAG 398

QY 3229 ggtcactgtggcctgtaacctgtctcttctgttgatcgctggcaacctggcacaggggg 3288
Db 399 GGCCACTCAGGGGCAATCACAGCTGTGTACATTGAT-----CAGACCATGGTACTGGCC 452

QY 3289 tctgggtcccgagcagcgctcctcgatggatcgtgttcacggagcgtcatgtat 3348
Db 453 AGTGGAGGACCAAGATGAGCCATCTGCCTGTGGATGTACTAACAGCAGCCGGGTACG 512

QY 3349 aatatacaagctcacgagcagcgctcagctgctggcctgtgtgcgccagttacgtaac 3408
Db 513 CAAACATTTGCTCACCGTGGAGATGTTACCTCCCTCACCTGTACCGCTTCTCTGTGCTATT 572

QY 3409 tcgctcggcagcagcagagagatttcggtatgggaacgatttcagggaacctgttgact 3468
Db 573 AGTAGTGTGCTGGATGACTTTCATCAGTATCTGGGACCGCAGCAGCATCAAGCTGTAC 632

QY 3469 accat 3473
Db 633 TCCAT 637

RESULT 9

AL043581
LOCUS AL043581 611 bp mRNA EST 29-FEB-2000
DEFINITION DKF2p434F0527_r1 434 (synonym: htes3) Homo sapiens cDNA clone DKF2p434F0527 5', mRNA sequence.

ACCESSION AL043581 GI:5422968
VERSION AL043581.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished (1999)
COMMENT On May 11, 1999 this sequence version replaced gi:4776630.
Contact: Bloecker H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
s1 sequence also available.
This clone (DKFZp434F0527) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

```

FEATURES
source
    Beijing Characterization, Genbank, Email: C10nest@zpu.de.
    Location/Qualifiers
        i. 6b1
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFP434F0527"
            /clone_lib="434 (synonym: htes3)"
            /tissue_type="testis"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Vector: pSport1; Site_1: Not1; Site_2: SalI"
BASE COUNT
    120 a 178 c 181 g 132 t
ORIGIN

```

Query Match	1.5%	Score 60.2;	DB 47;	Length 611;
Best Local Similarity	50.1%;	pred. No. le-06;		
Matches 180;	Conservative 0;	Mismatches 173;	Indels 6;	Gaps 1;
QY 3115	cagcagccaatcacatgcagtcgagtcgctaaacacacatggttttcactggcagccagat	3174		
Db				
1	CAAAAACCCATCACAGCCCTGAAAGCCGCTGCTGGGGCTTGGTACTGGGAGCAAGAC	60		
QY 3175	cacacctcaagtgattgcttccataaagtcggatgttgagtacagctccacgggtcac	3234		
Db				
61	CACACACTGAGAGTGTCCCGTCTGGAGACTCGTGCCTTTCACCCCTTCAGGGCCAC	120		
QY 3235	tgtgggctgtaaacctgtctctttgttgatcgctggaacctggcacagggggtcctg	3294		
Db				
121	TCAGGGCCATCACAGACCGTGTACATTAC-----CAGACCATGGTGCTGGCCAGTGA	174		
QY 3295	tcccagacgacctgtctcgtatgagatctgttcacggagacctgcatgataata	3354		
Db				
175	GGACAAGATGGGGCATCTGCCTGTGGATGTACTGTGACGTGCAGCCGGGTGAGCCATGTG	234		
QY 3355	caagctcacgacgagccgtacgtgcctggctgtgcccaggttacgtaatctcgcta	3414		
Db				
235	TTTGCTCACCGTGGGATGTACCTCCCTTACCTGTACCACCTCCTGTGTCATCAGCAGT	294		
QY 3415	ggcagcagcagagagatttgcgatgggaacgatattcagggaacacctgttgataccat	3473		
Db				
295	GGCCTGGATGACCTCATCAGCATCTGGGACCGCAGCAGCAGCATCAAGTTCTACTCCAT	353		

[illegible]

similar to cell division control protein CDC4, mRNA sequence.

AA354156
VERSION AA354156.1 GI:2006473
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georghiou,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P., Olsen,H., Raymond,L., Welj,Y.F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 12140200
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900162.
 Other_ESTs: THCL94741
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tadb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

```

seq primer: m13 reverse.
FEATURES
    source
        1..333
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="ATCC (inhost):154386"
                /db_xref="taxon:9606"
                /clone_lib="Jurkat T-cells v"
                /cell_type="T-lymphocyte"
                /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT      65 a      105 c      91 g      68 t      4 others
ORIGIN
Query Match      1.3%; Score 51.8; DB 25; Length 333;
Best Local Similarity 49.8%; Pred. No. 0.00028;
Matches 156; Conservative 0; Mismatches 151;
Indels 6; Gaps 1

```

Qy	3109	gccatcagcagcaatcacatgcatgcatgagctcggttaacgacatggttttcactggcagc	3168
Db	23	GCACCCAAAACCCATCACAGCCTGAAAGCGCTGTGGCGCTTGCTGACTGGGAGC	82
Qy	3169	caagatcacacccctcaaggctgattgcttcaataagatcggaatgttgagatatacagctccac	3228
Db	83	CAAGACCACACATGAGAGTNTTCCGTCTGGAGGACTCGTCTGCCCTTTCACCCCTTCAG	142
Qy	3229	ggtcactgtgggcctgataacctgtctctttgtggatcgtctggcaacctggcacagagggg	3288
Db	143	GGGCATCTAGGGGCCATCAGACCGCTNTACATTGAC-----CAGACCATGGTCTGCGCC	196
Qy	3289	tcctgattcccaagaagcgcgtctctgcatatagatctgctatcaggaagcctgcgatgat	3348

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:30:21 : Search time 183.89 Seconds
(without alignments)
3087.881 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgctgctgactgttttgta.....tgactaaatgttaaaataaaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCRTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.8	1.0	7218	1	US-08-232-463-14
2	39.4	1.0	7218	1	US-08-232-463-14
3	35.2	0.9	4195	2	US-08-340-011-1
4	35.2	0.9	4425	1	US-08-222-616-31
5	35.2	0.9	4425	6	PCT-US95-04228-31
6	35.2	0.9	4795	2	US-08-340-011-3
7	35.2	0.9	9108	6	PCT-US95-04228-45
8	34.4	0.8	1201	5	US-09-274-642-1
9	34.4	0.8	2375	5	US-08-714-918-81
10	34.4	0.8	2447	4	US-09-014-969-14
11	33.6	0.8	790	2	US-08-393-985-22
12	33.6	0.8	2097	2	US-08-393-985-1
13	33	0.8	10952	2	US-08-602-036A-1
14	33	0.8	10952	3	US-08-502-374A-1
15	33	0.8	10952	3	US-08-642-407A-1
16	32.8	0.8	1689	1	US-07-991-867B-41
17	32.8	0.8	1689	4	US-08-544-332-41
18	32.8	0.8	8457	1	US-07-991-867B-1
19	32.8	0.8	8457	4	US-08-544-332-1
20	32.4	0.8	12124	1	US-08-181-271A-36
21	32.4	0.8	12124	1	US-08-449-315-36
22	32.4	0.8	12124	1	US-08-444-803-36
23	32.4	0.8	12124	1	US-08-449-043-36
24	32.4	0.8	12124	2	US-08-456-265A-36
25	32.4	0.8	12124	2	US-08-455-416-36
26	32.4	0.8	12124	2	US-08-455-244-36

C 27	32.4	0.8	12124	2	US-08-454-876-36	Sequence 36, Appl
C 28	32.4	0.8	12124	3	US-08-457-364-36	Sequence 36, Appl
C 29	32.4	0.8	12124	3	US-08-456-262-36	Sequence 36, Appl
C 30	32.4	0.8	12124	3	US-08-455-736-36	Sequence 36, Appl
C 31	32.4	0.8	12124	4	US-08-971-217-36	Sequence 36, Appl
C 32	32.4	0.8	20303	1	US-08-370-975B-6	Sequence 6, Appl
C 33	32.4	0.8	26764	1	US-08-370-975B-1	Sequence 1, Appl
C 34	32.4	0.8	2338	2	US-08-425-069-1	Sequence 1, Appl
C 35	32.2	0.8	2338	4	US-08-317-844B-1	Sequence 1, Appl
C 36	32.2	0.8	4008	5	US-08-307-896-5	Sequence 5, Appl
C 37	32.2	0.8	4008	6	PCT-US95-11808-5	Sequence 5, Appl
C 38	32.2	0.8	1613	3	US-08-813-940-1	Sequence 1, Appl
C 39	31.6	0.8	7225	5	US-08-286-819A-15	Sequence 15, Appl
C 40	31.6	0.8	7225	5	US-08-980-357-15	Sequence 15, Appl
C 41	31.6	0.8	500	5	US-09-141-000-2	Sequence 2, Appl
C 42	31.4	0.8	2634	5	US-08-911-853-30	Sequence 30, Appl
C 43	31.4	0.8	17612	5	US-08-911-853-29	Sequence 29, Appl
C 44	31.4	0.8	1165	2	US-08-624-125-11	Sequence 11, Appl
C 45	31.2	0.8	1165	2	US-08-624-125-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14


```

; LOCATION: (610)..(942)
US-09-274-642-1

Query Match          0.8%; Score 34.4; DB 5; Length 1201;
Best Local Similarity 59.0%; Pred. No. 1.8;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 3535 aaatggglaagccctcattcttgatgccaataataagaggagacagtaataataaattt 3594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 AAATTCGGCAACTCCTCTCTCCCTTTATATCAGACTGAAAATTCAGCCAAATAATGCCATA 484

QY 3595 aattccaacacagagatctcttattgtgtggatgtgcga 3634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 ATTGCTAACCAATGACGCTCTCATTTGGATGGAGTGTGTGCA 444

RESULT 9
US-08-714-918-81/c
; Sequence 81, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,918
; FILING DATE: September 13, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/005
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-714-918-81

Query Match          0.8%; Score 34.4; DB 5; Length 2375;
Best Local Similarity 60.9%; Pred. No. 2.8;
```

```

Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1986 aaagctcctcttagcgatcgatataaaacgaacggcgagaaatgatgagacttcgata 2045
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1556 AAAAGCCCTTGAATCGTGCCATCTTCACCATTTAGGACCATGTATAATAATGGGAATACTGCA 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2046 tccagcttcgacgtacaaactatttcctttcaa 2077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1496 TCGTATGGTGTCTGCTGAACCTACTTCTTTTCAA 1465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-014-969-14

Query Match          0.8%; Score 34.4; DB 4; Length 2447;
Best Local Similarity 48.4%; Pred. No. 2.8;
Matches 92; Conservative 1; Mismatches 97; Indels 0; Gaps 0;

QY 3939 tgtacgtatctctctctagaattttgtgttagttgttgagacactaagttaacagctac 3998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2173 TATATGTTTCACATATCTGGATCTGTGTAATTTGTTACTTTTAATGTGACAAATAA 2232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3999 gctccggtagggttaaggaaactaaatgaatgaatataacacatatattttcgcg 4058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2233 ACCTTTTGGAGAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 2292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4059 taattataaaactacatagtgcttlaaagcgctcagcctcagcctataataataatgactaaatg 4118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
; METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,036A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: HYZ-039CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-602-036A-1

Query Match 0.8%; Score 33; DB 2; Length 10952;
Best Local Similarity 53.5%; Pred. No. 19;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 881 aaaagctgctgcagacactatcccccaactcccgcttggtgtagtgctgagagccga 940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5663 AACACAGCATTTTTCACCCCTCAGTCAGTCATCTATCCACTCCACAGCTCAGCCACAAA 5604

QY 941 cgaccataactacatctttttccaggagagtagtcagagtggtggagctggtgccttaca 1000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5603 GGCCCCGCCCTTGGCATTCCTTTCCCTCCAGAACTGGGTGCTGCTGGAGGCGCTGCC 5544

QY 1001 cagtggcct 1009
||| ||| |||
Db 5543 CAGGGGCGCT 5535

RESULT 14
US-08-374A-1/c
; Sequence 1, Application US/08502374A
; Patent No. 5872007
; GENERAL INFORMATION:
; APPLICANT: Fodstad, Oeystein
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL-SPECIFIC OLIGONUCLEOTIDES AND
; METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Hale and Dorr LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/502.374A
; FILING DATE: 14-Jul-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: HYZ-039DV2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-526-6000
; TELEFAX: 617-526-5000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-502-374A-1

Query Match 0.8%; Score 33; DB 3; Length 10952;
Best Local Similarity 53.5%; Pred. No. 19;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 881 aaaagctgctgcagacactatcccccaactcccgcttggtgtagtgctgagagccga 940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5663 AACACAGCATTTTTCACCCCTCAGTCAGTCATCTATCCACTCCACAGCTCAGCCACAAA 5604

QY 941 cgaccataactacatctttttccaggagagtagtcagagtggtggagctggtgccttaca 1000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5603 GGCCCCGCCCTTGGCATTCCTTTCCCTCCAGAACTGGGTGCTGCTGGAGGCGCTGCC 5544

QY 1001 cagtggcct 1009
||| ||| |||
Db 5543 CAGGGGCGCT 5535

RESULT 15
US-08-642-407A-1/c
; Sequence 1, Application US/08642407A
; Patent No. 5877308
; GENERAL INFORMATION:
; APPLICANT: Oeystein, Fodstad
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
; METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:39:04 : search time 299.13 seconds
(without alignments)
3452.656 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 ggtgcctgactgtttgtga.....tgactaaatgttataaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43.4	1.1	2481	1 X28627	Nucleotide sequenc
2	36.6	0.9	110000	1 V21209_14	Continuation (15 o
3	35.2	0.9	4195	1 T12068	FLT4 receptor tyro
4	35.2	0.9	4425	1 T03090	Protein tyrosine-k
5	35.2	0.9	9108	1 T03104	Plasmid pRK5.tk1-1
6	35	0.8	110000	1 V21209_07	Continuation (8 of
7	34.4	0.8	279	1 V86487	EST clone A111. N
8	34.4	0.8	2447	1 V54587	Human secretory pr
9	34.4	0.8	7363	1 V74330	Staphylococcus aur
10	34.4	0.8	237326	1 V57903	Hereditary haemoch
11	34.2	0.8	376	1 Q39893	Expressed Sequence
12	34.2	0.8	376	1 Q59305	Human brain expres
13	34	0.8	3060	1 X07378	Arabidopsis TTG1 g
14	34	0.8	3060	1 X07374	Arabidopsis TTG1 g
15	34	0.8	3060	1 X07376	Arabidopsis TTG1 g
16	34	0.8	3060	1 X07377	Arabidopsis TTG1 g
17	34	0.8	3060	1 X07379	Arabidopsis TTG1 g
18	34	0.8	3060	1 X07380	Arabidopsis TTG1 g
19	34	0.8	3060	1 X07383	Arabidopsis TTG1 g
20	34	0.8	28690	1 X13075	Enterococcus faeca
21	33.6	0.8	2097	1 V01546	Rat syntaxin 1A ge
22	33.6	0.8	7019	1 X13064	Enterococcus faeca
23	33.4	0.8	13104	1 Q46852	Clone of recombina
24	33	0.8	2522	1 V63196	cDNA from clone fp
25	33	0.8	2809	1 X13371	Enterococcus faeca
26	33	0.8	4239	1 V74644	Staphylococcus aur
27	33	0.8	10952	1 T33345	Human CAPL gene. S
28	33	0.8	10952	1 V41162	Human CAPL gene.
29	33	0.8	10952	1 X17696	Human CAPL genomic
30	32.8	0.8	1689	1 Q66824	AmEPV NPH-1 gene.
31	32.8	0.8	2118	1 V84592	Human secreted pro
32	32.8	0.8	8457	1 Q66797	AmEPV spheroidin g
33	32.6	0.8	507	1 V24841	H. pylori ORF hp4p

C	34	32.6	0.8	753	1	Q78153	Barley mild mosaic
C	35	32.6	0.8	1320	1	V25064	H. pylori cytoplasm
C	36	32.6	0.8	3485	1	Q53461	BAMV coat protein
C	37	32.6	0.8	7965	1	V74632	Staphylococcus aur
C	38	32.4	0.8	12124	1	V81625	Cucumber chitinase
C	39	32.4	0.8	12124	1	V62854	Cucumber chitinase
C	40	32.4	0.8	20303	1	T71699	Human deoxyeytidyl
C	41	32.4	0.8	26764	1	T71696	Human deoxyeytidyl
C	42	32.2	0.8	409	1	V88259	EST clone DY516. N
C	43	32.2	0.8	2338	1	Q14183	N. clavipes draglin
C	44	32.2	0.8	2338	1	V23249	Nephila clavipes s
C	45	32.2	0.8	4008	1	T14528	Rat adenyl cycla

ALIGNMENTS

RESULT	1
X28627	ID X28627 standard; cDNA; 2481 BP.
AC	X28627; AC
DT	07-JUN-1999 (first entry)
DE	Nucleotide sequence of Sel-10 gene.
KW	Nematode; SEL-10 gene; Alzheimer's disease; cancer; ss.
OS	Caenorhabditis elegans.
Key	Location/Qualifiers
FT	91..1854
FT	/*tag= a
FT	/product= "Sel-10"
FT	85..86
FT	/*tag= b
FT	/note= "SL1 splice junction"
FT	179..180
FT	/*tag= c
FT	/note= "SL1 splice junction"
FT	247..248
FT	/*tag= d
FT	/note= "SL1 splice junction"
FT	344..345
FT	/*tag= e
FT	/note= "SL1 splice junction"
FT	469..594
FT	/*tag= f
FT	/note= "F-Box"
FT	479..480
FT	/*tag= g
FT	/note= "SL1 splice junction"
FT	625..626
FT	/*tag= h
FT	/note= "SL1 splice junction"
FT	746..747
FT	/*tag= i
FT	/note= "SL1 splice junction"
FT	847..936
FT	/*tag= j
FT	/note= "WD40 (I)"
FT	969..1062
FT	/*tag= k
FT	/note= "WD40 (II)"
FT	1059
FT	/*tag= l
FT	/note= "optionally a G to A transition"
FT	1095..1182
FT	/*tag= m
FT	/note= "WD40 (III)"
FT	1136..1137
FT	/*tag= n
FT	/note= "SL1 splice junction"
FT	1215..1302
FT	/*tag= o
FT	/note= "WD40 (IV)"
FT	1335..1428
FT	/*tag= p


```

FT FT /*tag= b
FT FT /note= "these bases represent a line of missing text in
FT FT the sequence listing in the specification. They
FT FT are included to maintain the nucleotide numbering
FT FT given in the specification for this DNA sequence"
FT FT
FT FT misc_feature 4501..4560
FT FT /*tag= c
FT FT /note= "these bases represent a line of missing text in
FT FT the sequence listing in the specification. They
FT FT are included to maintain the nucleotide numbering
FT FT given in the specification for this DNA sequence"
FT FT
FT FT misc_feature 6301..6360
FT FT /*tag= d
FT FT /note= "these bases represent a line of missing text in
FT FT the sequence listing in the specification. They
FT FT are included to maintain the nucleotide numbering
FT FT given in the specification for this DNA sequence"
FT FT
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 267-271; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 7363 BP; 2558 A; 1013 C; 1286 G; 2262 T;

Query Match 0.8%; Score 34.4; DB 1; Length 7363;
Best Local Similarity 60.9%; Pred. No. 7.1;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1986 aaagctctctagcgtcgatataaaacgagccgaggaatgatgagacttcgata 2045
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1690 AAAGCCCTTGAAATCGTCGCTATTCACCATAGGACCATGTAATATGGGAATACTGCA 1631

Qy 2046 tccagcttcgacctaaactattctcttcaa 2077
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1630 TCGTATGCTGTCCTGGAACACTACTTCTTTTCAA 1599

RESULT 10
V57903/c
ID V57903 standard; DNA; 237326 BP.
AC V57903;
DT 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.

```

```

OS Homo sapiens.
PN WO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 01-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK.
DR WPI; 98-240014/21.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 1: Fig 9: 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BTF genes, which are homologues of the milk
CC protein butyrophilin (BT), and can be used in the production of agonists
CC and antagonists of BT function. Also described are: (1) a RoRet gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia.
SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

Query Match 0.8%; Score 34.4; DB 1; Length 237326;
Best Local Similarity 47.8%; Pred. No. 83;
Matches 130; Conservative 0; Mismatches 141; Indels 1; Gaps 1;

Qy 3858 ggaagtcattgacgcattcatttttatataaataatatacacattatagggtccgcagca 3917
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58136 GCAAAGGATAATTGAAATCCCAAACTTACAGGTTTTCACAAAGTAAAGTTTGCTAAA 58067

Qy 3918 acttcacggttttaacacaaagctgacgcatctcatctctagaaatttttgtagttgtg 3977
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58066 AGTTAACAGTGTAAACATGTATTATATAGTAACCTTCTAATCTTGTGGCCTTAGACAGTCTACT 58007

Qy 3978 gacactaagtgtacacgctacgctccggtaggtttaaggaactaaactaaatgaatcagat 4037
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58006 CCACAGACATAAAAGAAGTTCGCTTTGGAAAAGAAATGGTTATCATCTCCGGAATAAAAAA 57947

Qy 4038 atatacacatatatttttcgctgaattatataaacta-catagtgtcttaaacgcctcag 4096
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 57946 AAGGAATAAAGCAGGAGCGAGAAATTTATATAAAAAAGAAATGTTGTATCGAAAAATCCTTCT 57887

Qy 4097 cctaataaaaatgcactaaaatttataaaataaa 4128
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 57886 CCTGAGATAAATAACTAGTTGTTTAAAGAAA 57855

RESULT 11
Q39893/c
ID Q39893 standard; DNA; 376 BP.
AC Q39893;
DT 20-MAY-1993 (first entry)
DE Expressed Sequence Tag human gene marker EST00207.
KW Expressed sequence tag; human genome project; chromosome;
KW human gene sequencing; PCR mapping; somatic cell hybrids;
KW sublocalisation; gene tagging; tissue typing.
OS Synthetic.
PN WO9300353-A.
PD 07-JAN-1993.

```



```
Best Local Similarity 50.6%; Pred. No. 5.1;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 3963 ttgtgttagttgttgagactaaagtgttaacagctacgctcccgtaggttaagaagactaaa 4022
DB 2605 TTCTATATGTTTCTCAGGGGCTACAGCAGAGTCAGCAGTAAACAATTCFCAGATACTTCG 2546
QY 4023 ctaaatgaatcagatatacacatatattttcgcgtaattataataaacacatagtgtc 4082
DB 2545 TTAATTAACAATTTGTATACGTAGCTAGTGGCTTAAATTAAGAATAGATTGTC 2486
QY 4083 ttaagcgctcagcctaataataaaagactaaatgttaaaa 4124
DB 2485 TCACAAAAAATAGACTGAAGAAATTATGGGTAAAAATTAGAA 2444

RESULT 14
X07374/c
ID X07374 standard; DNA: 3060 BP.
AC X07374;
DT 07-JUN-1999 (first entry)
DE Arabidopsis TTG1 gene encoding plant regulatory protein.
KW TTG1 gene; transparent testa glabra 1 gene; signal transduction;
KW WD40 repeat protein; transgenic plant; trichome; anthocyanin;
KW pigment; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 1041..2075
FT mutation /*tag= a
FT FT
FT FT
FT FT
PN WO9900501-Al.
PD 07-JAN-1999.
PF 25-JUN-1998; G01861.
PR 25-JUN-1997; GB-013430.
PA (PLAN-) PLANT BIOSCIENCE LTD.
PI Gray JC, Walker AR;
DR WPI: 99-095746/08.
DR P-PSDB; W97835.
PT New isolated transparent testa glabra 1 gene - used for producing
PT plants with modified traits, e.g. insect protection, chemical or
PT fibre production, salt removal, pigmentation, taste or seedling
PT growth
PS Claim 3; Fig 3; 6lpp; English.
CC This is the nucleotide sequence of the novel Arabidopsis thaliana
CC Landsberg erecta TTG1 gene (transparent testa, glabra 1 gene),
CC which encodes a WD40 repeat protein (see W97835) suggested to be a
CC component of a signal transduction pathway which regulates
CC expression or action of downstream transcription factors. TTG1
CC is thought to act upstream of an Arabidopsis R homologue in the
CC pathways leading to trichome differentiation and anthocyanin
CC synthesis. The TTG1 locus was identified by complementation of a
CC ttg1 mutant with genomic DNA and by sequencing the gene in several
CC ttg1 mutant alleles. The 1.6 kb transcript is present in all major
CC organs. Mutant alleles (see X07376-80 and X07383) of TTG1 are also
CC claimed, as well as vectors, host cells, transgenic plants,
CC polypeptide expression products and other related materials, plus
CC methods of manipulating the phenotypic characteristics associated
CC with TTG1, e.g. the number of trichomes on the aerial parts of the
CC plant, number of trichomes on the root hairs, mucilage of the seeds,
CC dormancy of the seeds, anthocyanin pigmentation, condensation of
CC the tannins, number of stomata on hypocotyls. In particular,
CC insect protection, chemical production, climate tolerance, salt
CC removal, fibre production, ornamental value, water and nutrient
CC absorption, initiation of seed germination, pigmentation, taste, or
CC speed of seedling growth can be modified.
CC Sequence 3060 BP; 864 A; 600 C; 565 G; 1031 T;

Query Match 0.8%; Score 34; DB 1; Length 3060;
Best Local Similarity 50.6%; Pred. No. 5.1;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 3963 ttgtgttagttgttgagactaaagtgttaacagctacgctcccgtaggttaagaagactaaa 4022
DB 2605 TTCTATATGTTTCTCAGGGGCTACAGCAGAGTCAGCAGTAAACAATTCFCAGATACTTCG 2546
QY 4023 ctaaatgaatcagatatacacatatattttcgcgtaattataataaacacatagtgtc 4082
DB 2545 TTAATTAACAATTTGTATACGTAGCTAGTGGCTTAAATTAAGAATAGATTGTC 2486
QY 4083 ttaagcgctcagcctaataataaaagactaaatgttaaaa 4124
DB 2485 TCACAAAAAATAGACTGAAGAAATTATGGGTAAAAATTAGAA 2444
```

```
DB 2605 TTCTATATGTTTCTGAGGCGCATACAGCAGAGTCAGCAGTAAACAATTCFCAGATACTTCG 2546
QY 4023 ctaaatgaatcagatatacacatatattttcgcgtaattataataaacacatagtgtc 4082
DB 2545 TTAATTAACAATTTGTATACGTAGCTAGTGGCTTAAATTAAGAATAGATTGTC 2486
QY 4083 ttaagcgctcagcctaataataaaagactaaatgttaaaa 4124
DB 2485 TCACAAAAAATAGACTGAAGAAATTATGGGTAAAAATTAGAA 2444

RESULT 15
X07376/c
ID X07376 standard; DNA: 3060 BP.
AC X07376;
DT 07-JUN-1999 (first entry)
DE Arabidopsis TTG1 gene mutant allele ttg1.10.
KW TTG1 gene; transparent testa glabra 1 gene; signal transduction;
KW WD40 repeat protein; transgenic plant; trichome; anthocyanin;
KW pigment; mutant; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 1041..2075
FT mutation /*tag= a
FT FT
FT FT
FT FT
FT FT
PN WO9900501-Al.
PD 07-JAN-1999.
PF 25-JUN-1998; G01861.
PR 25-JUN-1997; GB-013430.
PA (PLAN-) PLANT BIOSCIENCE LTD.
PI Gray JC, Walker AR;
DR WPI: 99-095746/08.
DR P-PSDB; W97835.
PT New isolated transparent testa glabra 1 gene - used for producing
PT plants with modified traits, e.g. insect protection, chemical or
PT fibre production, salt removal, pigmentation, taste or seedling
PT growth
PS Claim 12; Page -: 6lpp; English.
CC This is the nucleotide sequence of the Arabidopsis thaliana
CC Landsberg erecta TTG1 gene (transparent testa, glabra 1 gene)
CC mutant allele ttg1.10. It contains a point mutation (G to A) in
CC the 5' untranslated region when compared to the wild-type TTG1
CC gene (see X07374). The gene encodes a WD40 repeat protein (see
CC W97835) suggested to be a component of a signal transduction
CC pathway that regulates expression or action of downstream
CC transcription factors. TTG1 is thought to act upstream of an
CC Arabidopsis R homologue in the pathways leading to trichome
CC differentiation and anthocyanin synthesis. The TTG1 gene, mutant
CC alleles (see X07376-80 and X07383), vectors, host cells, transgenic
CC plants, and polypeptide expression products are claimed, plus
CC methods of manipulating the phenotypic characteristics associated
CC with TTG1, e.g. the number of trichomes on the aerial parts of the
CC plant, number of trichomes on the root hairs, mucilage of the seeds,
CC dormancy of the seeds, anthocyanin pigmentation, condensation of
CC the tannins, number of stomata on hypocotyls. In particular,
CC insect protection, chemical production, climate tolerance, salt
CC removal, fibre production, ornamental value, water and nutrient
CC absorption, initiation of seed germination, pigmentation, taste, or
CC speed of seedling growth can be modified. The mutation found in
CC ttg1.10 may affect the translation of RNA. Plants having the
CC ttg1.10 allele have more trichomes, different seed mucilage and
CC less anthocyanin.
CC (NB. the nucleotide sequence for allele ttg1.10 was produced from
CC information provided in Fig 3 of the specification).
CC Sequence 3060 BP; 865 A; 600 C; 564 G; 1031 T;

Query Match 0.8%; Score 34; DB 1; Length 3060;
Best Local Similarity 50.6%; Pred. No. 5.1;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

QY 3963 ttgtgttagtttggacactaagtgtacacagctacgctccggtagggttaaggaactaaa 4022
Db 2605 TTCTATATGTTTCTGAGGGCATACACGAGAGTCAGCAGTAACAATTCTCAGATACTTCG 2546
QY 4023 ctaaatgaatcagatatacacatatatttttcgogtaattatataaaactacatagtc 4082
Db 2545 TTAATAATAACAATATTGTATACGTAGCTAGTGGCTTCGTTAAATTAGAATAGATTGTC 2486
QY 4083 ttaaagcgctcagccctaataataaaatgactaaaatgttaaaa 4124
Db 2485 TCAAAAAAAAAATAGACTGAAGAATTATGGGTAAAAATTAGAA 2444

Search completed: September 4, 2000, 00:58:07
Job time: 18062 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:29:28 ; Search time 183.89 Seconds
(without alignments)
1474.373 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcagcgaggttaatgctg.....aaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
7: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	2.3	7218	1	US-08-232-463-14
2	38.6	2.0	2007	5	US-08-747-221B-36
3	38.6	2.0	2007	5	US-08-747-221B-38
4	38.2	1.9	3238	6	PCT-US94-10080-5
5	38	1.9	936	4	US-08-557-309B-21
6	38	1.9	936	5	US-08-834-306-21
7	37.6	1.9	7218	1	US-08-232-463-14
8	37	1.9	684	2	US-08-226-264-27
9	36.6	1.9	314	2	US-08-686-878A-25
10	36.4	1.8	1454	3	US-08-713-000-7
11	36.4	1.8	1454	4	US-08-975-316-7
12	36.4	1.8	1474	4	US-08-975-316-71
13	36.4	1.8	7859	1	US-07-854-596B-4
14	36.4	1.8	7859	3	US-08-450-905B-15
15	36.4	1.8	7859	5	US-07-982-759F-15
16	35.2	1.8	19124	4	US-08-487-826B-13
17	35	1.8	946	5	US-08-916-443A-15
18	34.8	1.8	972	1	US-07-915-934-1
19	34.8	1.8	972	1	US-08-325-743-1
20	34.8	1.8	1023	1	US-08-252-966B-16
21	34.8	1.8	1223	5	US-09-154-874-4
22	34.6	1.8	536	1	US-08-341-568-1
23	34.6	1.8	536	3	US-08-911-020-1
24	34.4	1.7	1307	4	US-08-960-022-17
25	34.4	1.7	1639	4	US-08-737-524B-1
26	34.4	1.7	1897	1	US-08-184-632-1

27 34.4 1.7 2143 3 US-08-656-177A-1 Sequence 1, Appli
28 34 1.7 1776 5 US-08-655-352-10 Sequence 10, Appl
29 34 1.7 2901 6 PCT-US96-00419-4 Sequence 4, Appli
30 33.8 1.7 940 3 US-08-471-717-1 Sequence 1, Appli
31 33.8 1.7 1602 2 US-08-530-950-3 Sequence 3, Appli
32 33.8 1.7 2026 4 US-08-993-228-3 Sequence 3, Appli
33 33.8 1.7 4843 5 US-08-986-485-1 Sequence 1, Appli
34 33.6 1.7 1393 1 US-08-174-467-18 Sequence 18, Appli
35 33.6 1.7 1393 5 US-08-452-071-18 Sequence 18, Appli
36 33.6 1.7 1395 4 US-08-553-367A-1 Sequence 1, Appli
37 33.6 1.7 1792 5 US-08-747-221B-18 Sequence 18, Appli
38 33.6 1.7 1792 5 US-08-747-221B-20 Sequence 20, Appli
39 33.6 1.7 2550 7 5258287-23 Patent No. 5258287
40 33.6 1.7 3527 4 US-08-909-965C-7 Sequence 7, Appli
41 33.4 1.7 1687 1 US-08-143-219-26 Sequence 26, Appl
42 33.4 1.7 1813 6 PCT-US94-12883-3 Sequence 3, Appli
43 33.4 1.7 1898 1 US-08-342-411A-1 Sequence 1, Appli
44 33.2 1.7 1137 2 US-08-706-214-2 Sequence 2, Appli
45 33.2 1.7 1804 4 US-08-504-459-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEJFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/834,306
;; FILING DATE: 15-APR-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.422C1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 21:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 936 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-834-306-21

Query Match 1.9%; Score 38; DB 5; Length 936;
Best Local Similarity 62.8%; Pred. No. 0.1;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 1877 tgcctagtccttagcttaactcgtatttttttgcgtggttcgagctcaaaaaa 1936
Db 841 TGTGAAGCTCTTGTCTTTTCTTTCTATTTTGAACGGTGATTCGCATATATA 900

QY 1937 aaataaataagctcaaaaaa 1970
Db 901 TTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 934

RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: Foley & Lardner
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7218 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; CLONE: ptz9pt-FLS
;; US-08-232-463-14

Query Match 1.9%; Score 37.6; DB 1; Length 7218;
Best Local Similarity 4.4%; Pred. No. 0.47;
Matches 16; Conservative 191; Mismatches 155; Indels 0; Gaps 0;
QY 916 aacacgtagtgtcactgaactaacaagaaatccccgcgtgagggagagcgcgcttgc 975
Db 1396 RRR 1337
QY 976 aagtgacaatacaataaaccgaagtaaacgctgccacgtataaacagcagagagttgg 1035
Db 1336 RRR 1277
QY 1036 tgacatgctgcagaactcttgaagctcaagccggctactgtgtgagtgogactcg 1095
Db 1276 RRR 1217
QY 1096 tgcagcttaacagcagaagcagcgccatctcacatcatagcattggtcagctacagt 1155
Db 1216 RRR 1157
QY 1156 gctgtgatgaactaaatccgacgttaagctgtcttcgaggtggaggagcgaatggag 1215
Db 1156 RRR 1097
QY 1216 atgtgccggtgagctgcgcgagcagtgatgtctcaatgtgcgcgcaactttggagagg 1275
Db 1096 RRR 1037
QY 1276 tc 1277
Db 1036 TC 1035

RESULT 8
US-08-226-264-27
; Sequence 27, Application US/08226264
; Patent No. 5801017
; GENERAL INFORMATION:
; APPLICANT: Werber, Moshe M.
; APPLICANT: Zeelon, Elisha P.
; APPLICANT: Levanon, Avigdor
; APPLICANT: Guy, Rachel
; APPLICANT: Goldlust, Arie
; APPLICANT: Rigbi, Meir
; APPLICANT: Panet, Amos
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: PRODUCTION OF RECOMBINANT FACTOR XA
; TITLE OF INVENTION: INHIBITORS OF LEECH HIRUDO MEDICINALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; CITY: 1185 Avenue of the Americas
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/226,264


```

QY 1929 caaaataaaatcaaatgaactaaataaaaaa 1970
||||| |||| ||| || ||| ||||| |||||
Db 1411 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1452

RESULT 11
US-08-975-316-7
; Sequence 7, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; NAME: SLEATH, Janet
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-975-316-7

Query Match 1.8%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 59.8%; Pred. No. 0.41;
Matches 61; Conservative 0; Mismatches 41; Indels 0;
Gaps 0;

QY 1869 tagcttttgcataagcttaactgaactgtatttttcgtacggtgtcagct 1928
||||| |||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1351 TAAATAATATCCAAATGCTTCGCCAAATTAATATGTTATGTTATATGAAA 1410

QY 1929 caaaataaaatcaaatgaactaaataaaaaa 1970
||||| |||| ||| || ||| ||||| |||||
Db 1411 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1452

RESULT 12
US-08-975-316-71
; Sequence 71, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; NAME: SLEATH, Janet
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-975-316-71

Query Match 1.8%; Score 36.4; DB 4; Length 1474;
Best Local Similarity 59.8%; Pred. No. 0.41;
Matches 61; Conservative 0; Mismatches 41; Indels 0;

QY 1869 tagcttttgcataagcttaactgaactgtatttttcgtacggtgtcagct 1928
||||| |||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1371 TAAATAATATCCAAATGCTTCGCCAAATTAATATGTTATGTTATGTTATAT 1474

QY 1929 caaaataaaatcaaatgaactaaataaaaaa 1970
||||| |||| ||| || ||| ||||| |||||
Db 1431 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1472

RESULT 13
US-07-854-596B-4
; Sequence 4, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplowski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0 Version #1.25

```



```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/854,596B
; FILING DATE: 03-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 92,337
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..7859
; OTHER INFORMATION: /note= "sequence of plasmid pSW6"
US-07-854-596B-4
```

```
Query Match 1.8%; Score 36.4; DB 1; Length 7859;
Best Local Similarity 66.7%; Pred. No. 1.1; Mismatches 0; Gaps 0;
Matches 52; Conservative 0; Indels 0;

Qy 1893 cttaactcgtattatttcgtacggtgtcgcagtcacaaatacaaatcaagctca 1952
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5575 CTTAAAGAGATTCTCTTTTATGATATTTGTACAAAAA 5634
```

```
Qy 1953 aaaaaa 1970
Db 5635 AAAAAA 5652
```

```
RESULT 14
US-08-450-905B-15
; Sequence 15, Application US/08450905B
; Patent No. 5856301
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Stem Cell Inhibiting Proteins
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.905B
; FILING DATE: 26-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/982,759
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127319.3
; FILING DATE: 23-DEC-1991
; APPLICATION NUMBER: GB 9221587.0
; FILING DATE: 14-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, HOLLIE L.
; REGISTRATION NUMBER: 31,321
```

```
; REFERENCE/DOCKET NUMBER: 102.378.120DV-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-526-6110
; TELEFAX: 617-526-5000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-450-905B-15
```

```
Query Match 1.8%; Score 36.4; DB 3; Length 7859;
Best Local Similarity 66.7%; Pred. No. 1.1; Mismatches 0; Gaps 0;
Matches 52; Conservative 0; Indels 0;

Qy 1893 cttaactcgtattatttcgtacggtgtcgcagtcacaaatacaaatcaagctca 1952
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5575 CTTAAAGAGATTCTCTTTTATGATATTTGTACAAAAA 5634
```

```
Qy 1953 aaaaaa 1970
Db 5635 AAAAAA 5652
```

```
RESULT 15
US-07-982-759F-15
; Sequence 15, Application US/07982759F
; Patent No. 6057123
; GENERAL INFORMATION:
; APPLICANT: CRAIG, Stewart
; APPLICANT: GEORGE, Michael
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: CZAPLEWSKI, Lloyd George
; APPLICANT: GILBERT, Richard
; TITLE OF INVENTION: Stem Cell Inhibiting Proteins
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,759F
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127319.3
; FILING DATE: 23-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9221587.0
; FILING DATE: 14-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, HOLLIE L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 102378.120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-526-6000
; TELEFAX: 617-526-5000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
```

US-07-982-759F-15

Query Match 1.8%; Score 36.4; DB 5; Length 7859;
Best Local Similarity 66.7%; Pred. No. 1.1;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 1893 cttaactcgtatttttttcgtacggttgctcagctcaaaaaataaaatcaaaattaaagcta 1952
DB 5575 CTTAAAGATTCTCTTTTTTTTATGATATTTGTACAAAAA
QY 1953 aaaaaaaaaaaaaaa 1970
DB 5635 AAAAAAAAAAAAAA 5652

Search completed: September 4, 2000, 00:30:21
Job time: 1991 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	41	2.1	1651	1	Q80218	Human NDF-alpha2b
2	41	2.1	2335	1	O80216	Human proNDF-alpha
3	40.2	2.0	1586	1	Q47967	Rape acyl-ACP thio
4	39.6	2.0	1078	1	Q25785	Protein synthesis
5	38.6	2.0	2007	1	V40754	C. felis esterase,
6	38.6	2.0	2007	1	V40755	C. felis esterase,
7	38.2	1.9	3238	1	Q83532	Sequence encoding
8	38	1.9	936	1	T69171	Trypanosoma cruzi
9	37.8	1.9	1378	1	V59706	Human secreted pro
10	37.8	1.9	2323	1	V59524	Human secreted pro
11	37	1.9	684	1	Q72956	Clone PSP65-Xai-11
12	36.6	1.9	314	1	Q09108	3' nucleotide sequ
13	36.6	1.9	1118	1	Q30263	Deg-1 gene, DNA se
14	36.6	1.9	2116	1	Q35987	Tomato hsp80 cDNA
15	36.6	1.9	3628	1	N60848	Plasmid sequence e
16	36.4	1.8	1454	1	V23913	Plant CAD enzyme D
17	36.4	1.8	1474	1	V23914	Plant CAD enzyme D
18	36.4	1.8	2353	1	V45981	A. thaliana sterol
19	36.4	1.8	7859	1	Q12154	Shuttle vector pSW
20	36.4	1.8	7859	1	O44265	PSW6 for expressio
21	36.4	1.8	7984	1	Q25185	PSW6 expression ve
22	36	1.8	1285	1	Q86744	TIMP-3 metalloprot
23	36	1.8	2218	1	H84503	Human secreted pro
24	36	1.8	3828	1	V06566	Arabidopsis cellul
25	36	1.8	4358	1	V42653	Nucleotide sequen
26	35.6	1.8	465	1	V34173	Human secreted pro
27	35.6	1.8	1099	1	V34238	Human secreted pro
28	35.6	1.8	1410	1	Q15022	Hyoscyanine 6 beta
29	35.6	1.8	10660	1	Q84793	Spinocerebellar at
30	35.4	1.8	106	1	V00430	3', fragment of clo
31	35.4	1.8	801	1	N91234	cDNA clone 8F3 of
32	35.2	1.8	1203	1	T64815	Tumour suppressor
33	35.2	1.8	19124	1	T72882	Plasmodium var-7 q

PT Pathogen resistant plants prodm. using protein synthesis inhibiting
PT gene - or its fusion products, inserted into the genome, also
PT pharmaceutical use of derived synthesis inhibiting protein
PS Disclosure, Fig 3; 23pp; German.
CC The sequence is that of a gene encoding a protein which effectively
CC blocks protein synthesis by plant pathogens, e.g. Trichoderma reesi
CC and Fusarium sporotrichoides. It can be fused to active promoters
CC such as wun-1 and the fusion introduced into plant genetic material
CC to impart pathogen resistance to both mono- and di-cotyledonous
CC plants. The gene was isolated from mature barley seeds.
SQ Sequence 1078 BP; 276 A; 310 C; 321 G; 171 T;

Query Match 2.0%; Score 39.6; DB 1; Length 1078;
Best Local Similarity 63.8%; Pred. No. 0.19;
Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1877 tgcatagtccttagcttaactcgattatttctgacggttcgagctcaaaaata 1936
Db 979 TGTCAAGAGTTACATATAAACAATAATAATAATTAATAATGTCACAGTTTAAAAAAA 1038

Qy 1937 aaatcaattaaagctcaaaaaaataaaaaa 1970
Db 1039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1072

RESULT 5
V40754
ID V40754 standard; cDNA; 2007 BP.
AC V40754;
DT 23-SEP-1998 (first entry)
DE C. felis esterase, nFE92007, coding sequence.
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE92007; ds.
OS Ctenocephalides felis.
FH Key Location/Qualifiers
FT CDS 11..1597
FT CD5 /*tag= a
FN WO9821324-A1.
PD 22-MAY-1998.
PF 10-NOV-1997; U20598.
PR 12-NOV-1996; US-747221.
PI (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
DR WPI: 98-297929/26.
P-PSDB: W57862.
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
PS Claim 1; Page 160-162; 230pp; English.
CC This sequence encodes the flea esterase protein, nFE92007 (the
CC complementary strand is shown in V40755), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in development,
CC metamorphosis, feeding, digestion and reproduction.
SQ Sequence 2007 BP; 695 A; 286 C; 371 G; 655 T;

Query Match 2.0%; Score 38.6; DB 1; Length 2007;
Best Local Similarity 61.4%; Pred. No. 0.49;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1870 agcttttgcatagtcccttagcttaactcgattatttctgacggttcgagctc 1929
Db 101 ATCTTGTATCAATATTCGCTTTATTTTTCATTTTTTTTCAATAATATATGTTT 42

Qy 1930 aaaaataaaatcaaatgaagctcaaaaaaataaaaaa 1970
Db 41 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1

RESULT 7
Q83532
ID Q83532 standard; DNA; 3238 BP.
AC Q83532;
DT 28-SEP-1995 (first entry)
DE Sequence encoding serine/threonine kinase receptor CFK1-10a.
KW Bone morphogenic protein; receptor; serine/threonine kinase; BMP;
KW bone; cartilage; injury; treatment; inhibition; ss.
OS Rattus rattus.
FH Key Location/Qualifiers

Qy 1870 agcttttgcatagtcccttagcttaactcgattatttctgacggttcgagctc 1929
Db 1907 ATCTTGTATCAATATTCGCTTTATTTTTCATTTTTTTTCAATAATATATGTTT 1966

Qy 1930 aaaaataaaatcaaatgaagctcaaaaaaataaaaaa 1970
Db 1967 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2007

RESULT 6
V40755/c
ID V40755 standard; cDNA; 2007 BP.
AC V40755;
DT 23-SEP-1998 (first entry)
DE C. felis esterase, nFE92007, coding sequence complementary strand.
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE92007; ds.
OS Ctenocephalides felis.
PN WO9821324-A1.
PD 22-MAY-1998.
PF 10-NOV-1997; U20598.
PR 12-NOV-1996; US-747221.
PI (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
DR WPI: 98-297929/26.
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
PS Claim 1; Page 165; 230pp; English.
CC This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE92007 (see V40754 for coding strand), of the
CC invention. When administered to animals, the protein induces a protective
CC immune (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in development,
CC metamorphosis, feeding, digestion and reproduction.
SQ Sequence 2007 BP; 655 A; 371 C; 286 G; 695 T;

Query Match 2.0%; Score 38.6; DB 1; Length 2007;
Best Local Similarity 61.4%; Pred. No. 0.49;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1870 agcttttgcatagtcccttagcttaactcgattatttctgacggttcgagctc 1929
Db 101 ATCTTGTATCAATATTCGCTTTATTTTTCATTTTTTTTCAATAATATATGTTT 42

Qy 1930 aaaaataaaatcaaatgaagctcaaaaaaataaaaaa 1970
Db 41 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1

RESULT 7
Q83532
ID Q83532 standard; DNA; 3238 BP.
AC Q83532;
DT 28-SEP-1995 (first entry)
DE Sequence encoding serine/threonine kinase receptor CFK1-10a.
KW Bone morphogenic protein; receptor; serine/threonine kinase; BMP;
KW bone; cartilage; injury; treatment; inhibition; ss.
OS Rattus rattus.
FH Key Location/Qualifiers


```

PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057661.
PR 12-SEP-1997; US-057761.
PR 12-SEP-1997; US-057855.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
PI WPI; 98-506364/43.
DR P-PSDB; W74923.
DR New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders
Claim 1: Page 434-435: 721pp: English.
This sequence represents a nucleic acid molecule designated Gene 14 from
the human cDNA clone HPWFD84 (deposited as clone AFCC 97897 and AFCC
209043) which encodes a secreted human protein. The gene can be used to
generate fusion proteins by linking to the gene to a human immunoglobulin
Fc portion (e.g. V59502) for increasing the stability of the fused
protein as compared to the human protein only.
The invention relates to 186 novel genes and their fragments (nucleic
acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
are useful for preventing, treating or ameliorating medical conditions
e.g. by protein or gene therapy. Also, pathological conditions can be
diagnosed by determining the amount of the new polypeptides in a sample
or by determining the presence of mutations in the new polynucleotides.
Specific uses are described for each of the 186 polynucleotides, based on
which tissues they are most highly expressed in (see V59511 for described
uses).
Sequence 1378 BP; 494 A; 254 C; 248 G; 377 T;

Query Match 1.9%; Score 37.8; DB 1; Length 1378;
Best Local Similarity 69.9%; Pred. No. 0.68; Mismatches 22; Indels 0; Gaps 0;
Matches 51; Conservative 0;

Qy 1898 ctcgtatttatttcgtacggtgttcgagctcaaaaataaaatcaagctaaaaa 1957
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1263 CTGCTTCTTATTTAATAAATTATCAGAGTCAGAAAAAATAAAAAA 1322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1958 aaaaaa 1970
      ||||| ||||| |||||
Db 1323 AAAAAAAAAA 1335

RESULT 10
V59524
ID V59524 standard; DNA; 2323 BP.
AC V59524;
DT 06-JAN-1999 (first entry)
DE Human secreted protein gene 14 clone HPWFD84.
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN W09839448-A2.
PD 11-SEP-1998.
PF 06-MAR-1998; U04493.
PR 02-OCT-1997; US-061060.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
```

```
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.

PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PR (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI: 98-506364/43.
DR P-PSDB: W74744.
PR New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 252-253; 721pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 14 from
CC the human cDNA clone HPMFD84 (deposited as clone ATCC 97897 and ATCC
CC 209043) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W4731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 2323 BP; 760 A; 467 C; 438 G; 658 T;

Query Match 1.9%; Score 37.8; DB 1; Length 2323;
Best Local Similarity 69.9%; Pred. No. 0.87; Mismatches 0; Gaps 0;
Matches 51; Conservative 0; Indels 22;

QY 1898 ctcgattatttttcgtacgttgcagctcaaaataaataatcaaatcaagctaaaaa 1957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2201 CTGTGTTCTTTATTAATAATTTATCAGATGCAAAAAAATAAAAAAATAAAAAA 2260

QY 1958 aaaaaaataaaaaa 1970
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2261 AAAAAAATAAAAAA 2273

RESULT 11
Q72956
ID Q72956 standard; cDNA; 684 BP.
AC Q72956;
DT 04-JUN-1995 (first entry)
DE Clone pSP65-XaI-11 encoding recombinant Factor Xa inhibitor (FXaI).
KW Leech; factor Xa inhibitor; ss.
OS Hirudo medicinalis.
FH Key Location/Qualifiers
FT cds 139..684 /*tag= a
FT misc_difference 1..684 /*tag= b
FT /*label= N= unspecified base
PN W09423709-A.
PD 27-OCT-1994.
PF 08-APR-1994; U03918.
PR 09-APR-1993; US-045804.
PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.
PA (YISS ) YISSUM RES & DEV CO.
PI Fischer M, Goldlust A, Guy R, Levanon A, Panet A;
PI Rigbi M, Werber MM, Zeelon EP;
DR WPI: 94-341457/42.
DR P-PSDB: R62619.
PT Recombinant factor 10a inhibitor of Hirudo medicinalis - for
PT treating excessive blood coagulation, partic. thrombosis, also
PT related DNA, vectors, transformed cells and antibodies
```



```
cds
3. .1865
/*taq= a
```

US5187267-A.
16-FEB-1993.
19-JUN-1990; 541883.
19-JUN-1990; US-541883.
(CALJ) CALGENE INC.
Comai L, Koning AJ;
WPI: 93-075789/09.
P-PSDB: R32548.

New DNA constructs - comprising tomato heat shock protein 80 gene non-coding region and gene region of interest, useful for producing transgenic plants with modified phenotype(s)
Claim 2: Fig 1: 33bp: English.

The sequence given encodes the tomato heat shock protein 80.5 (hsp80). Tomato hsp80 is characterised by having a mRNA of about 2.3 kb, two introns, a predicted pI of about 4.69 and a molecular weight of 80,479.8 daltons. The polypeptides encoded by tomato hsp80 have substantial homology with the polypeptides encoded by related heat shock proteins from a variety of prokaryotic and eukaryotic species. There are two hsp80 genes in tomato, identified as locus A and locus B. The low copy number of the hsp80 gene is a useful indicator of the strength of the promoter. Tomato hsp80 message is found in abundance in most, if not all meristematic tissue eg. flower meristems, early fruit, and root and shoot apices. hsp80 message is detected in floral primordia of immature fruit, where it increases until the fruit begins to swell. The mRNA levels observed in normal mature leaves is approx. 1/10th of the level observed in meristematic tissue.

Sequence 2116 BP; 647 A; 349 C; 562 G; 558 T;

Query Match
1.9%; Score 36.6; DB 1; Length 2116;

Best Local Similarity 60.6%; Pred. No. 1.8;

Matches	60;	Conservative	0;	Mismatches	39;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Qy 1872 cttttgtcatagtccttagtcttaactcgtatttatttcgtacgggtgtcgagctcaa 1931

11

Oy 1932 aaataaaatcaaatataaqaataaaaaaaaaa 1970

Address	Disassembly	Comment
00000000	CALL 00000000	
00000001	CALL 00000000	
00000002	CALL 00000000	
00000003	CALL 00000000	
00000004	CALL 00000000	
00000005	CALL 00000000	
00000006	CALL 00000000	
00000007	CALL 00000000	
00000008	CALL 00000000	
00000009	CALL 00000000	
0000000A	CALL 00000000	
0000000B	CALL 00000000	
0000000C	CALL 00000000	
0000000D	CALL 00000000	
0000000E	CALL 00000000	
0000000F	CALL 00000000	
00000010	CALL 00000000	
00000011	CALL 00000000	
00000012	CALL 00000000	
00000013	CALL 00000000	
00000014	CALL 00000000	
00000015	CALL 00000000	
00000016	CALL 00000000	
00000017	CALL 00000000	
00000018	CALL 00000000	
00000019	CALL 00000000	
0000001A	CALL 00000000	
0000001B	CALL 00000000	
0000001C	CALL 00000000	
0000001D	CALL 00000000	
0000001E	CALL 00000000	
0000001F	CALL 00000000	
00000020	CALL 00000000	
00000021	CALL 00000000	
00000022	CALL 00000000	
00000023	CALL 00000000	
00000024	CALL 00000000	
00000025	CALL 00000000	
00000026	CALL 00000000	
00000027	CALL 00000000	
00000028	CALL 00000000	
00000029	CALL 00000000	
0000002A	CALL 00000000	
0000002B	CALL 00000000	
0000002C	CALL 00000000	
0000002D	CALL 00000000	
0000002E	CALL 00000000	
0000002F	CALL 00000000	
00000030	CALL 00000000	
00000031	CALL 00000000	
00000032	CALL 00000000	
00000033	CALL 00000000	
00000034	CALL 00000000	
00000035	CALL 00000000	
00000036	CALL 00000000	
00000037	CALL 00000000	
00000038	CALL 00000000	
00000039	CALL 00000000	
0000003A	CALL 00000000	
0000003B	CALL 00000000	
0000003C	CALL 00000000	
0000003D	CALL 00000000	
0000003E	CALL 00000000	
0000003F	CALL 00000000	
00000040	CALL 00000000	
00000041	CALL 00000000	
00000042	CALL 00000000	
00000043	CALL 00000000	
00000044	CALL 00000000	
00000045	CALL 00000000	
00000046	CALL 00000000	
00000047	CALL 00000000	
00000048	CALL 00000000	
00000049	CALL 00000000	
0000004A	CALL 00000000	
0000004B	CALL 00000000	
0000004C	CALL 00000000	
0000004D	CALL 00000000	
0000004E	CALL 00000000	
0000004F	CALL 00000000	
00000050	CALL 00000000	
00000051	CALL 00000000	
00000052	CALL 00000000	
00000053	CALL 00000000	
00000054	CALL 00000000	
00000055	CALL 00000000	
00000056	CALL 00000000	
00000057	CALL 00000000	
00000058	CALL 00000000	
00000059	CALL 00000000	
0000005A	CALL 00000000	
0000005B	CALL 00000000	
0000005C	CALL 00000000	
0000005D	CALL 00000000	
0000005E	CALL 00000000	
0000005F	CALL 00000000	
00000060		

RESULT 15

KE30LI
N60848

ID N60848 standard; DNA; 3628 BP.

AC N60848;
ID N60848

DT 24-OCT-1991 (first entry)

DE Plasmid sequence encoding bovine pre-prolactin hormone.

KW *E. coli*; growth hormone; bGH; bPRL; ds.

Bos taurus.

PN J61202691-A.

PD 08-SEP-1986

PF 04-MAR-1985; 042406.

PR 04-MAR-1985; JP-042406.

PA (NAKA/) NAKAJIMA K.

DR WPI; 86-276389/42.

Plasmid(s) contg. ampicillin resistant genes - for amplification of bovine growth hormone and prolactin genes.

PS Disclosure; Table 1-18; 13pp; Japanese.

The bovine pre-prolactin hormone encoding sequence is derived from

CC pituitary tissue. The plasmids carry an amp resistance gene and

CC and may be expressed by an *E. coli* host in an ampicillin medium

CC for efficient and large scale production of the hormone.

SQ	Sequence	3628 BP;	965 A;	873 C;	901 G;	889 T;

Query Match

Best Local Similarity 65.18; Pred. No. 2.3;

Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1888 ttagtcttaactcgtatttattttcgtacggttgcgagctcaaaaataaaaatcaaatta 1947

```

Db      1183  TTCATCTTTAAATAAACACAGACTCTGTGACGATGTCAAAATCTAAAAA  

QY      1948  agctaaaaaaiaaaaaaaaaa 1970  

Db      1243  AAAAAAAAAAAAAAAAAAAAAA 1265

```

Search completed: September 4, 2000, 00:39:04
Job time: 16919 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 23:42:35 ; Search time 11551.9 Seconds
(without alignments)
304.508 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgagattaatgctg.....aaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: em_fun:*
- 17: em_humi:*
- 18: em_hum2:*
- 19: em_in:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_sy:*
- 29: em_un:*
- 30: em_vi:*
- 31: gb_htg1:*
- 32: gb_htg2:*
- 33: gb_in1:*
- 34: gb_in2:*
- 35: em_ba1:*
- 36: em_ba2:*
- 37: em_hum3:*
- 38: em_hum4:*
- 39: gb_pr4:*
- 40: gb_htg3:*
- 41: gb_htg4:*
- 42: gb_htg5:*
- 43: gb_htg6:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_pl3:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1806.4	91.6	77862	34	AC005465	AC005465 Drosophil
2	1806.4	91.6	78927	43	AC020317	AC020317 Drosophil
3	1806.4	91.6	249001	58	AE003824	AE003824 Drosophil
4	88.6	4.5	1732	12	AF019611	AF019611 Cricetulu
5	86	4.4	1759	11	AF019612	AF019612 Homo sapi
6	46.4	2.4	3946	33	DDU25143	U25143 Dictyosteli
7	46	2.3	7218	5	I56494	I56494 Sequence 14
8	45	2.3	108908	33	PFMAL3P8	AL034560 Plasmodiu
9	44.6	2.3	173184	60	AC013583	AC013583 Homo sapi
10	44.6	2.3	237302	40	AL136310	AL136310 Homo sapi
11	44.2	2.2	268	33	DMTNF14B	X01937 Drosophila
12	44.2	2.2	17196	34	AE003083	AE003083 Drosophil
13	44.2	2.2	24781	34	AE003098	AE003098 Drosophil
14	44.2	2.2	51236	41	AC006589	AC006589 Drosophil
15	44.2	2.2	61267	34	AC004336	AC004336 Drosophil
16	44.2	2.2	82692	19	AC003924	AC003924 Drosophil
17	44.2	2.2	100729	41	AC008141	AC008141 Drosophil
18	44.2	2.2	104494	51	AC012166	AC012166 Drosophil
19	44.2	2.2	129241	51	AC010113	AC010113 Drosophil
20	44.2	2.2	139238	19	AC006302	AC006302 Drosophil
21	44.2	2.2	139561	55	AC007904	AC007904 Drosophil
22	44.2	2.2	174982	54	AC013424	AC013424 Drosophil
23	44.2	2.2	176878	42	AC006091	AC006091 Drosophil
24	44.2	2.2	198282	61	AC007415	AC007415 Drosophil

25	44.2	2.2	224263	34	AE003720	Drosophil
c	26	44.2	2.2	224263	34	AE003720
27	44.2	2.2	225419	34	AE003703	Drosophil
c	28	44.2	2.2	229655	34	AE003707
c	29	44.2	2.2	230977	34	AE003729
30	44.2	2.2	251154	34	AE003656	Drosophil
31	44.2	2.2	253172	34	AE003718	Drosophil
32	44.2	2.2	254846	34	AE003725	Drosophil
33	44.2	2.2	257672	34	AE003517	Drosophil
c	34	44.2	2.2	261000	34	AE003654
c	35	44.2	2.2	267271	58	AE003800
c	36	44.2	2.2	303885	34	DROSADH01
c	37	44.2	2.2	316735	58	AE003786
c	38	43.2	2.2	93240	42	AC007586
c	39	43.2	2.2	134837	55	AC007888
c	40	43.2	2.2	179122	34	AC006414
c	41	43.2	2.2	224939	34	AE003710
c	42	42.6	2.2	1546	34	AE003035
c	43	42.6	2.2	40653	51	AC011906
c	44	42.6	2.2	122539	51	AC010110
c	45	42.6	2.2	127091	53	AC009843
RESULT	1					
AC005465						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						

FEATURES						
Source						
Location/Qualifiers						
1. 77862						
/organism="Drosophila melanogaster"						
/strain="y2; cn bw sp"						
/db_xref="taxon:7227"						
/chromosome="2R"						
/map="48C1-48C2"						
/clone.lib="p1 library, partial Sau3A in pAdiosacBII"						
/clone="P1 DS05372 (D379)"						
/note="Between bp 20,516 and bp 20,517, Tn10 and the 9 bp duplication were excised."						
BASE COUNT	21405 a	16792 c	16981 g	22684 t		
ORIGIN						
Query Match		91.6%	Score 1806.4;	DB 34;	Length 77862;	
Best Local Similarity		94.2%	Pred. No. 0;			
Matches 1945;	Conservative	0;	Mismatches	1;	Indels	118; Gaps
2;						
QY	10	gattaatgctgattcttgcgtgactcacacagcattgctggtataagagagtcggagacca	69			
Db	13013	GATTAAATGCTGATTTCTGCTGGACTACACAGCATTGTGGTATTAAGAGTTCGGGACCA	13072			
QY	70	gaggagtaagatttcgggaaggaatcccggtccggtagggaactactagcatttcgcaagtga	129			
Db	13073	GAGGAGTAAGATTTCGGGAAGGAATCCCGTCGGTAGGGACTACTAGCATTCCCAAGTGA	13132			
QY	130	qgtccagacaacggaggaccctcccaactgtagaatacgcatacaccatctctaaatcccaacaa	189			
Db	13133	CGTCCAGCAACCGGAGGACCCCAACTGTAGTAATCCGATACCATCTCTTAATCCCAACAA	13192			
QY	190	accaatgacatcttgagacctcaacgacgatggtaccccttcgttcttctcatagactgag	249			
Db	13193	ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATGACTCTGG	13252			
QY	250	catgctttatggcgctttcttacttttgcacgcttttca-----	291			
Db	13253	CATCGCTTATGGCGTCTCTTTACTTTTTCGACCGCTTCTTCAAGGTGTAGTATATCCAGC	13312			
QY	292	-----	-----			
Db	13313	CAAGTACGTCAGATACTTAATGTAAATCCCTTGTAGTTGTCATGCACCTACCCGTACGATG	13372			
QY	319	cttctcctcaagaacacccgggctgagtataaaattcatgacctccactggcacacagagtg	378			
Db	13373	CCTTCTCAAGAACACCGGGCTGAGTATAAATTTTCATGAGCCTCCACTGGCACACGAGTG	13432			
QY	379	octttaacaggaccctcctaogctgggagatctgcgggttaacagctgcacccggagagtaa	438			
Db	13433	CCTTTAACAGGACCCCTCCTACGCTGGGATCTGCGGTAAACAGCTGCACCCGGAGAGTAA	13492			
QY	439	tgatcacacagctttaatgtaggagtcctggtaaccttttctctgctccgatcgctctga	498			
Db	13493	TGATCACCAGCTTTAATGTAGGAGTCTCGGTGTCACCTTTCTCTGCCGATCGGTTCTGA	13552			
QY	499	tctctctattgctccactatcttcacagcagtggtgaaacagacagctcttctgctglatcct	558			
Db	13553	TCCTGCTCATTCGCCACTATCTTCACAGTGGTGAACAGACAGCTCTTCGTCTGTATCCT	13612			
QY	559	gcgccgttgagtcctctgacagctggaaatcttactgcgcggcgctcaacttcgcgttgg	618			
Db	13613	CGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTTCTACTGCCCGCGCTCAACTTCCGCTGG	13672			
QY	619	agggatcggactactacatcacacaccttctgctctgctgctggtggtgacgagatggagac	678			
Db	13673	AGGATCGGATACATACATCACAAACCCCTTGCTCTGCTTGGTGTGACGAGATGGGAC	13732			
QY	679	acgcctggcgcgctgtagtggaggatgtgcctgtcacccgggttttggaaataa-----	729			
Db	13733	ACGCCCTGGCCCTGTGATGGAGGATGTGCCTGTACCGGGTTTGGAAATAAAGGTTAGAA	13792			
QY	730	-----	-----			

Db	13793	TGATATAGCGTACTTTACAACTAATACGCTTACTAAATTAGGTAAATAATTAATATTTCAG	13855
Qy	732	ttcattcttgcctgccgttagcaTaccagagactctcccagcaccacttaaacagctcta	791
Db	13853	TTCATCTTCCGCTGCGCTTAGCATACAGGAGCTCTCCACGAGCCACTTAACAGACTCA	13912
Qy	792	cgttggtcccaagctacgtgttctgtgcgctggaactctggcataatttggttgcgt	851
Db	13913	CGTTGGTTCCGCAAGCTACGTTCTGTGTGCGCTGGAATCTGGCAATAATTTGTGTTCGCT	13972
Qy	852	ggcgtgactatctcttaactcaacggtgggaatcactatgtaacacttttgtacgcttaac	911
Db	13973	GGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCACTTTGTACGCTTAC	14032
Qy	912	aaccaacgctagtggtcactgaaactaaaggaataccccgcctgagggagagcgcgcg	971
Db	14033	AACCAACACGTAGTGGTCACTGAACTAACAGGAAATCCCCGCTCAGGGGAGAGCGCGC	14092
Qy	972	ttgcaagtggacaatacacaataacccaadcaaacggctgccagtgcccagtaaacagcgagagagt	1031
Db	14093	TTGCAAGTGGACAATCAAAATACCCCAAGTAACCGGCTGCCACGTAACACAGCGAGGAGT	14152
Qy	1032	tgggtgacatgccctgcagaaactctctaagctcaagccggcactgctgtgagtgcggac	1091
Db	14153	TGGGTGACATGCCCTGCAGAACTCTCTGAAGCTCTCAAGCCGGGCTACTGTGTGAGTCCGGAC	14212
Qy	1092	ttcgtgcagcttaacgagaaagcagcgccactctcaatcatagcattgtaTggtcaagcta	1151
Db	14213	TTCTGTGCAGTTAACGACGAAGCAGCGCCATCTCACATCATAGCATGTGATGGTCACTCA	14272
Qy	1152	cagtgcgtgatgaactaaataccgaaagttaagctgcttcgaggttggtggaggaacgcaaat	1211
Db	14273	CAGTCTGTGATGAACATAAATCCGAACCTAAAGCTGCTTCGAGGTGGTGGAGCAGCAAAAT	14332
Qy	1212	ggagaTgTgcggtggagactccgcagcatatgtctcaatgtgcgccgcactttggag	1271
Db	14333	GGAGATGTGCGGTGGAGCTGCCGACGATGTATGTCTCAATGTGCGCGGCACATTTGGAG	14392
Qy	1272	gaggtctccgagcactgctctccggagtttgcaacagggattctgctcagcaccgtt	1331
Db	14393	GAGGTCTCCGAGCACTGCTCGCTCCGGAGTTTGCAACAGAGGATTTCTGCCTACGACCGCTT	14452
Qy	1332	atcagaaataTcactgcaataatgacttcaagcgacagaatatttcgcgagagaaactg	1391
Db	14453	ATACGAAATATCACTGCCATTAATACGTTCAAGCGACAGAAATTTTCGCGGAGAGAAAGCTG	14512
Qy	1392	ccgcgctgactctatgtgggccatccatggagatgtaactcgaaactgtgagaagtatccgc	1451
Db	14513	CCGCCGCTGATCTATFTGTGGCCATCCATFGGATGTCACTCCAAGTGTGGAGGTATCCGCC	14572
Qy	1452	tttgtgcgaagataTacttataaaggcagcttgccggagatgcttgactgctgctctctc	1511
Db	14573	TTTGTGCGAGATATAGCTTATTAAAGGCAGCTTGCGCGATGCCGCTGCTGCTCTCTC	14632
Qy	1512	aagtatacgttggtcttcaagcataggatggcgttgatcaatgcacattccccctgctttggt	1571
Db	14633	AAGTATAACGTTGGTCTTTCAGCATAGGATTGGCTTGTATCAATGCAATTCCTCTCTTT	14692
Qy	1572	ttcagtgggcccaattaccagcacgtgatacaacagctctcttggtgggcagagatggat	1631
Db	14693	TTCCATGCGCCCACTATPACACAGACCGCTGATACAGACTTCTTGTGTGGGACAGTGCAT	14752
Qy	1632	cagatgcgaagagagatatcatctcgttgataatcaccagcgaggttccctctctcttt	1691
Db	14753	CAGCATGCCAAGAGAGATATCATCTCGTTGTGATATCACACGCTGGGTTCCTCTCTCTT	14812
Qy	1692	gcactggccctgcttaagtgggcttggttagttttcttgcaaccctgctttaagaactg	1751
Db	14813	GCATGGCCCTCTTAAGGTGGCCTGGTTGAGTTTCTGTGCAACCCCTGCTTTAAGAACTG	14872
Qy	1752	aaatggaaaactgaaatggatctctggaggttcaactcccctgcgaagacgcttagactcta	1811
Db	14873	AAATGGAAAACCTGAAATGGATTCCTGGGAGTTCAACTCCTCTGCAAGACGCTPAGACTGCTA	14932

QY	1812	tttacccttcagaaacacacaaacacacgcgaattgttagcacctcaagagattcgatag	1871
Db	14933	TTTTACCTTTCACGAAACACACAAAACACACGGAATTTGTAGCACCTCAAGATTTCGATAG	14992
QY	1872	ctttttgcatagtccttagctcttaactcgtatttatttcttcatcagcttgcagctcaa	1931
Db	14993	CTTTTGTGTCATAGTCCCTTAGTCTTAACTCGPAITATTTTCGTACGGTTGTCGAGCTCAA	15052
QY	1932	aaataaaatcaaatcaagctaaaa	1955
Db	15053	AAATAAAATCAAAATTAAGCTAAGA	15076
RESULT	2		
AC020317			
LOCUS	AC020317	78927 bp	DNA
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***	HTG	03-JAN-2000
ACCESSION	AC020317		
VERSION	AC020317.1	GI:5664580	
KEYWORDS	HTG; HTGS_PHASE2.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;		
AUTHORS	Pterygota: Neoptera: Endopterygota: Diptera; Brachycera;		
TITLE	Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.		
JOURNAL	1 (bases 1 to 78927)		
	Adams,M. and Venter,J.C.		
	Direct Submission		
	Submitted (30-DPC-1999) Celera Genomics, 45 West Gude Drive,		
	Rockville, MD, USA		
COMMENT	This sequence was identified as CDM:10212749 by the submitter.		
	For more information on this record e-mail to fly@celera.com.		
	* NOTE: This is a 'working draft' sequence.		
	* This sequence will be replaced		
	* by the finished sequence as soon as it is available and		
	* the accession number will be preserved.		
FEATURES	Location/Qualifiers		
source	1..78927		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
BASE COUNT	22005 a 17623 c 17956 g 21343 t		
ORIGIN			
Query Match	91.6%	Score 1806.4;	DB 43; Length 78927;
Best Local Similarity	94.2%	Pred. No. 0;	
Matches 1945;	Conservative 0;	Mismatches 1;	Indels 118; Gaps 2;
QY	10	gattaatactgaatttcgtctgaggaagaaatcccgctcggtgaggaactaagcattccgaagtga	129
Db	52202	GATTAATGCTGATTTCTGGTCTGGACATACACAGATTTCCTGGTATAGGAGTCCGGACCA	52261
QY	70	gaggagtaagatttcgggaaggaatcccgctcggtgaggaactaagcattccgaagtga	129
Db	52262	GAGGAGTAAGATTTCGGAGAGGAATCCCGTCCGCTACGGACTACTAGCATTCGCAAGTGA	52321
QY	130	cgtcagcaacccggagaccccccaactgtagaatcccgcatcacatcctaatcccaacaa	189
Db	52322	CGTCAGCAACCGGAGGACCCCAACTGTAGATCCGCATCACCATCTTAATCCCAACAA	52381
QY	190	accaatgacatctllgagaccttcaaccacccatggaatcccttcglttcttcataagctgg	249
Db	52382	ACCAATGACATCTTTGAGACCTCACACGCCATGGATGCCCTTCGTTCATAGTACTGG	52441
QY	250	catcgcttatggcgcttcttacttttccagcgcttcttca-----	291
Db	52442	CATCGCTTATTCGGCTCTCTTACTTTTTCGACCGCTCTTCTTCAAGGTGCTACTATATCCAGC	52501
QY	292	-----agagttgcatgcactaccgctacgatg	318
Db	52502	CAAAGTAGTCCAGATTACTTAATGATTCCTTTAGCTTGCATGCCTACCTACCGTACCATG	52561

QY	319	cttctctcaagaacacccggctgagtgataaaatttccatgagctccacttgccacacagtg	378	QY	1332	atacgaataatcactgcccataatgacgttccagcagcagacaatttttcgcggagagaagctg	1391
Db	52562	CCTTCTCTAAGAACACCGGCTGAGTATAAATTTCTATGAGCCTCCACTGGCACACGAGTG	52621	Db	53642	ATACGAAATATCAC TGCCATAATGACGTTCAAGCGACAGAAATTTTCGCGGAGAGAGCTG	53701
QY	379	cttttaacaggaccctcctacgcgtggggtatcgcgggttaacagctgcacccggagagataa	438	QY	1392	ccgcgcgtgatctatgtgggccatccatgggatgtcaactcgaactgtgaggtatccgcc	1451
Db	52622	CCTTTAACAGGACCCCTCCTACCGTGGGATCTGCCGTAACAGCTGCACCCGGAGAGTAA	52681	Db	53702	CGCCCGTGATCTATGTGGGCCATCCATGGATGTCACTCGAACTGTGAGGTATCCGCC	53761
QY	439	tgatacagcgtttaatgttaggagtccttggtcaacettttctctctgctcccgatcgtctga	498	QY	1452	tttgtccagagatatagcttattaaaggcagcctgcccagatgcttgctgctgctcctc	1511
Db	52682	TGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTCTCTGCTCCCGATCGGTCCTGA	52741	Db	53762	TTTGTGCCGAGATATAGCTTATAAAGGCAGCCTGCCGAGATCCCTGCTGCTTCCTC	53821
QY	499	tctgctcattgpcactatcttcagcagtggtgaacaagacagctcttgcgtctgtatcct	558	QY	1512	aagataaacgctggtcttcagcatagattggcgttgaatcaatgccattccctcctttggt	1571
Db	52742	TCTGCTCATTTGCCACTATCTTCAGCAGTGGTGGAACAACAGACCTCTTCTGCTGTATCCT	52801	Db	53822	AAGTATAACGTGGTCTTCAGCATAGGATTGGCGTTGATCAATGCCATTCCCTGCTTTGGT	53881
QY	559	cgcccgctggagtcctctgtagcagctggaattctactgcccggcgtcaacttgcgcttg	618	QY	1572	ttcgatggcgcacattaccagcaccgtgatcacacagcttcttgggtggcagagtgagat	1631
Db	52802	CGCCCGTTGGAGTCCCTGTGCAGCTGGAATTTCTACTGCCCGCGCTCAACTTGCCTTGG	52861	Db	53882	TTTCGATGGCGCCACATTACCAGCACCGTCAACACAGCTTCTTGTGGGCGAGAGTGGAT	53941
QY	619	aggagatcggtactacatacaaaccttctgctctgcttgggtgacagagatgggac	678	QY	1632	cagcatgccaaagagatatcatctcgtttgataataaccagcgtgggttccctctctctt	1691
Db	52862	AGGAGATCGGATACTACATCACAAACCTTTGCTCTGCTTGGTGGTGCACGAGATGGGAC	52921	Db	53942	CAGCATGCCAAGAGAGATATCATCTCTGTTGATTAATCACCAGCGTGGGTTCCTTCTCTTT	54001
QY	679	acgccctggccgctgtgtaggagatgtccctgtccacgggttttggaataa	729	QY	1692	gcactggccctgcttaaggtggcctggttgagtttctgcgacccctgctttaagaactg	1751
Db	52922	ACGCCCTGCCGCTGTGATGGAGATGTGCCGTGTCACCGGGTTTGAATTAAGGTTAGAA	52981	Db	54002	GCAC TGGCCCTGCTTAAGGTGGCTGTGTTGAGTTTCTGCGACCCCTGCTTTAAGAACTG	54061
QY	730	-----	-ag 731	QY	1752	aaatggaaaactgaaatggatcctctgggagttcaactccctgaaaagacgctagactgcta	1811
Db	52982	TGATATAGCGCTACTTTTACAACCTAATACGCTTACTAATAGGTAAATATTAATATTTACG	53041	Db	54062	AAATGGAAAACCTGAATGGATCCTGGGAGTTCAACTCCCTGCAAGACGCTAGACTGCTA	54121
QY	732	ttoactcttgcctgcgttagcatcacacgagagctctccacgacacacttaaacagtcta	791	QY	1812	tttcaccttcacgaacacacacacacagcgaattgtagcacctcacaagattcagatag	1871
Db	53042	TTCATCTTCTGCCCTGCGGTAGCATACCGAGCTCTCCACACGACCCTTAACACGCTCA	53101	Db	54122	TTTCACCTTCAGAAACACACACACACAGCGAATTTAGCACCTCAAGATTCAGATAG	54181
QY	792	cgttggttccgaagctagctgttctgtgcgtggaaactctggcataattttgtgttcgct	851	QY	1872	cttttgcctagctccttagcttcaactcgtattttatcttcgacggttgcagctcaa	1931
Db	53102	CGTTGTTCCGCAAGCTACGTGTTCTGTCGCTGGAATCTGGCATAAATTTGTGTTCCGT	53161	Db	54182	CTTTTGTCTATAGTCCTTAGTCTTAACCTGATTTATTTTTCGTACGTTGTCGAGCTCAA	54241
QY	852	ggcgtgtgctatcttatactcaacggtgggaactcaactatgtcaacttgttacgcttac	911	QY	1932	aaataaaatcaaatgaactaaaa	1955
Db	53162	GGCGTGTGTTATCTTAATCTCAACGGTGGGAATCACTATGTATGATCAGCTTAC	53221	Db	54242	AAATAAAATCAAAATTAAGCTAAGA	54265
QY	912	aaccaacagctagtgtcactaaactaacaaggaaatccccctgagggagacgcgc	971	RESULT	3		
Db	53222	AACCAACAGCTAGTGTGTACTGAACATAACAGGAATATCCCGCTGAGGGAGAGCGCGC	53281	AE003824	249001 bp	DNA	INV 22-MAR-2000
QY	972	ttgcaagtggacaatacaatacccccaagtaaacggtgcctgaataacacgcgagagagt	1031	LOCUS	AE003824	Drosophila melanogaster genomic scaffold 142000013386047	section 34
Db	53282	TTGCAAGTGGACAATCAATAACCCCAAGTAAACGGCTGCCAGTAAACAGCGAGAGT	53341	DEFINITION	of 52, complete sequence.		
QY	1032	tgggtgacatgcctgcagaactctctgaagctcaaacggcggtcactgtgtgagtgccgac	1091	ACCESSION	AE003824	AE002787	
Db	53342	TGGGTGACATGCTGCGAAGTCTCTGAAGCTCAAGCGGGCTACTGTGTGAGTGGCGAC	53401	VERSION	AE003824.1	GI:7303542	
QY	1092	ttcgtcagcttaacagaaacagcgcacatctcaatcatagattgaattggtcagcta	1151	KEYWORDS	HTG		
Db	53402	TTCGTGCAGCTTAACGACGAAGACGCGCATCTCAATCATATAGCATATGATGTGATGCT	53461	SOURCE	fruit fly.		
QY	1152	cagtgtgtgataaactcaaatccgaacgtgaagctgtctcagagtggtgagagacgcaaat	1211	ORGANISM	Drosophila melanogaster		
Db	53462	CAGTGTGTGATCACTAAATCCGACGTAAGCTGCTTCGAGTGGTGGAGCGCAAT	53521	REFERENCE	1 (bases 1 to 249001)		
QY	1212	ggagatgtcccggtgagctgcgcagcagcatgtatgtctcaatgtgcgcgcaactttggag	1271	AUTHORS	Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Ammanides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galie, R.F., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bereman, B.P., Bhandari, D., Bolshakov, S., Borkov, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Meys, A.D., Dew, I., Dietz, S.M.,		
QY	1272	gaagttccagacactgctcgtccgagtttgcaacgagggattctgcctacgacccgctt	1331				
Db	53582	GAGGTCTCCGAGCACTGCTGCTCCGGAGTTTGCAACGAGGATTTGCTTACGACCGCTT	53641				

Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S.,
 Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S.,
 Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrelli, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Muzny, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
 Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 249001)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 FEATURES
 source
 Location/Qualifiers
 1..249001
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 complement(join(<5140..5357,5490..6855,6912..7040,
 7095..7172))
 /gene="CG8298"
 /product="CT24549"
 /db_xref="FLYBASE:FBan0008298"
 /db_xref="FLYBASE:FBgn0033673"
 /evidence=not_experimental
 complement(<5140..>7172)
 /gene="CG8298"
 /db_xref="FLYBASE:FBan0008298"
 /db_xref="FLYBASE:FBgn0033673"
 /evidence=not_experimental
 complement(join(5140..5357,5490..6855,6912..7040,
 7095..7172))
 /gene="CG8298"
 /note="CG8298 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008298"
 /db_xref="FLYBASE:FBgn0033673"
 /evidence=not_experimental
 /protein_id="AAF58597.1"
 /db_xref="GI:7303543"
 translation="MSNGRYGFGALIGVACVSSGTHCRFLIYTKNAEVLAYHEKLR
 QIVHAGMEYDPSEIWKNTQCIETAYKNLIVLEINPRDIIAIGINORTISVLNWL
 ETGQPLNAGWSDCRSPILKTLIHVRHNVYVRVRSGLPLSCPSALKIRLMDH
 VPVATAENKCLFGLDLSWLLNLTGGVGVHSTDTINAHRTSLMNVSTQWDPR
 LCQFRPLPLTPRINSEIFGYVLEGPUGHTPIAAMMGQPSALLGOLCVKAGQNV
 CLTDDSCFVLNTGREKLDSANGLITIAHKLGEKAATNYTLEGAISNAGSTVTLWRD
 KLOINTEINSNDNVESLNTFIGNSMISSCSMLNAECGLAAKSEITFVPAFHG
 MYAPYWRHDAIGIILGTSTAEITQAAVEATGFOIFEVLOAFKRDTPNWRDSSMO
 PVLTEGDAYENLHVOFIADIIGYMLERPOTTSPAGLGVMITAGVTMKVYVLEHAYK
 MYTPDPTVSPTTKNRELlyKKWATAYAKCLHWNFTYEADIELFAQRELPNLP
 VRRSVPGSIFLITFSFAMILIANYLKSPLK"
 join(7688..7799,7859..8060,8126..8420,8921..9141,
 9196..9395,9460..>9550)
 /gene="CG8972"

/product="CT25796"
 /db_xref="FLYBASE:FBan0008972"
 /db_xref="FLYBASE:FBgn0033672"
 <7688..>9550
 /gene="CG8972"
 /db_xref="FLYBASE:FBan0008972"
 /db_xref="FLYBASE:FBgn0033672"
 join(7753..7799,7859..8060,8126..8420,8921..9141,
 9196..9395,9460..9550)
 /gene="CG8972"
 /note="CG8972 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008972"
 /db_xref="FLYBASE:FBgn0033672"
 /protein_id="AAF58598.1"
 /db_xref="GI:7303544"
 translation="MLMSRALECRSLWLPQVARRCHANVVPILRINSHPAARSCRQIH
 SNRKSSNLKPTTGEPAAEQNTPPVNNVTKAVATGFTVGCFCAGATILLEYETRS
 LILFKARQARFCWQSRSLADRDYWTQIKODIRRHWSLTIPGDKMFAPILLCNLVAF
 MVRYPALKSTMTVTSTNPAKVVCWPMFLSTFESHYSAMHLFANNMYVHSPANAASV
 LGKEQLAVLYLSAGVSESLMSVLYKAATSOAGMSLGASGALMTLLAYVCTOYPTDOLS
 ILFLPALIFSAGAGIKVLVMDIDFAGVVMGWKFFDHAHLGGAMFGIFWATYGAQIWA
 RIGLLNYHDLRTRTKQK"
 join(11156..11206,11355..11603,12112..>12993)
 /gene="RnrS"
 /product="CT25800"
 /db_xref="FLYBASE:FBan0008975"
 /db_xref="FLYBASE:FBgn0011704"
 /evidence=not_experimental
 <11156..>12993
 /gene="RnrS"
 /note="CG8975"
 /map="48D-48D"
 /db_xref="FLYBASE:FBan0008975"
 /db_xref="FLYBASE:FBgn0011704"
 /evidence=not_experimental
 join(11186..11206,11355..11603,12112..>12993)
 /gene="RnrS"
 /note="RnrS gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008975"
 /db_xref="FLYBASE:FBgn0011704"
 /evidence=not_experimental
 /protein_id="AAF58599.1"
 /db_xref="GI:7303545"
 translation="WEKSLSPSKKILTDSTNNYRKMSICHEANGQLAKESSTVNGI
 GKSNLSMEKSVTPDPSLEPLLRNPRRFVIFQYIHDIMOMYKKAESPWVEED
 LSKDLTDWIRLKDDEHFIHVLAFFAASDGVINENLVERFQEQVQITEARCFYQFI
 AMENVHSEMYSVLIDTYIRDPHQREYLFNAIETMPAVKRKADWALSWSKSNFGER
 IIAFAAVEGIFFGSGFASIPWLKKRGLMPGLTFSNELISRDEGLHCDFAVLMFOHLVQ
 PKRERIIEIRDAVAIRQEFITDALPVNLIGMNCDEMSOYIEFVADRLLVELGVKGI
 YNTKNPFNFMEMISLDGKTNFFERKVGVEYRMGVVSNFLDNVFTLDADF"
 join(<13698..13986,14051..15157,15956..16090,16149..17250,
 17502..17536,17592..>18694)
 /gene="CG8976"
 /product="CT25806"
 /db_xref="FLYBASE:FBan0008976"
 /db_xref="FLYBASE:FBgn0033671"
 /evidence=not_experimental
 <13698..>18694
 /gene="CG8976"
 /db_xref="FLYBASE:FBan0008976"
 /db_xref="FLYBASE:FBgn0033671"
 /evidence=not_experimental
 /note="CG8976 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008976"
 /db_xref="FLYBASE:FBgn0033671"
 /evidence=not_experimental
 /protein_id="AAF58600.1"

```
/db_xref= "GI:7303546"
/translation= "MDRWNPKEKIEOSLSLLAAEYSSSSGAAASGEDETTTSGALR
SKALKRQLRNRLPMLRLGCTLSAFILGLLVLVPLVLYLDVHKRSLGPDW
TSETSRSDADYLDIGLSSGVIPKDFCRNKTFLVAVCTGVDFNFIORHTIRETWGNTT
EFNYPAFKLHGHUKHYLPLPDLKMYDYLSEGGOSLTASVRIVRIVGRQKDEAM
LGNETYLRHIESEKNDIIQENFVDSYNNLTLSVMALKHISRCSPNTAYFLKDD
DTFVNPILLNPLFLGGTIPLYNDYLDYHDSRSTYLVAPQRLKASSDLYLYGHQFNVV
PVSEVSMYPSYMYKESYKYLSDGAGYLSIDVQRLFEASLNTTFLVLEQVYIT
GLCAQKINRHHPELFSFAHSCOMKAFGRTIHOHLKDDSMVSAWNVSNYSKCP
PGRYSQVRLKRRPIYELKRLYRIVRIWLFRYKKAIIILAVFLAYIRFNKEDPSK
RTSLAUGSRDPRVADYLDPTKDTALIVRPHFRSKALLITIAVCSVHHFERRAI
RKLWGFNTDFNYSVFKLHGLKRGYQDLPRLKLYSEYLSGEGDSLRASIRLVFI
GRNLASLENEAIAEAQKNDVIQENFIDITYNNLTIKAVMKLHIQTSOLINTAFY
FKCDDDTFVNVPNLIHLFGGTIPVNYTAGFVGTNYEVTSPRKLRTAREMYGRQ
HCNVPATNKLNKMYPSYMPFGRGVYPRYLOGSGYLLSIDVVPRLYKASLTIRLVHLE
DMFVTGLCAERAGIKRTNHLPRSSYPYEGDEQCALKGSFTVHRAKDNVMEAYRVT
NFSKCPQPPQDLHURLKMQSPRPVDPVHDKRAELSGWEREKPLKIGDYLDQGNKA
LIVPRDFCKSALLIVAVCSGLDHFQESAIROTWNKTSFNFGFVRLHGLHGYL
SVMRGLKLYSMYLSGLDSDLSITAKIRIVIFILGRKNDKSKRELDKLFRESIOHNDIQE
NFVDSYHNLITKYSVMALKHISQSCADRAAFLFKCDDDTFVNVPNLIHLFGGTIPLYK
DTVGYSRTTYKVSQPNRLNGSRGLMYGHKFCNMTKVDVVKSPWYMPYMYFKGAKYP
KYLSTGTGYSIMSDIVKRLYAEALTTSLVHLEDVFTVGICAKKAGIRRRHQDLFNVHG
KPLCIFKGTITMHPVPLHSDMDAFVSNICLUPPRKHFEVKTLSKSHC"
complement(join(19346..19752,19879..20106,20173..21066,
21358..21831,21914..22864,23382..24697,24816..25284,
25347..>25744))
/seqname="CG8290"
/product="CT24533"
```

Query Match 91.6%; Score 1806.4; DB 58; Length 249001;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 1945; Conservative 0; Mismatches 1; Indels 118; Gaps 2;

```
QY 10 gattaatgctgatttctgttgactacacagcattgctggtataaaggagtcgggacca 69
Db 152260 GAFTAAATGCTGATTTCTGTGCTGTGACTACACAGCATTCGTGTATTAAGAGTCGGGACCA 152319
QY 70 gaggagtaagatttcgggaagaaatccctccggttaggactactagcatttcgcaagtga 129
Db 152320 GAGGAGTAAGATTCGGGAAGAAATCCCGTCGGTAGGGACTACTAGCATTCGCAAGTCA 152379
QY 130 cgtccgaacacggaggaccccccaactgtagaatccgcataccatccctaatacccaacaa 189
Db 152380 CGTCCAGCAACCGGAGAGCCCCCACTAGATAATCGGCATCACCATCTTAATCCCAACAA 152439
QY 190 acaatgacattctgaagacctcaccagccatgagatcccttgatttcttcatagctag 249
Db 152440 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTCTTCATAGTACTGG 152499
QY 250 catcgctttatggcgttcttacttttcgacgcgttcttca----- 291
Db 152500 CATCGCTTTATGGCGTCTTTACTTTTCGACCGCTCTCTTCAAGGTGTAGTATATCCAGC 152559
QY 292 -----agattgcatgactaccgcgtacgtg 318
Db 152560 CAAAGTACGTCCAGATACTTAATGTAATCCCTTAGAGTTGTCATGTGCATACCCGTACGATG 152619
QY 319 ccttctcgaacacacggcgtgagtatataatttcatgagcctccactggcacagagtg 378
Db 152620 CCTTCTCAAGAACACCGGGCTGAGTATATAATTTCTAGAGCTTCCACTGGCACACGAGTG 152679
QY 379 cctttaacaggaccctccacgtgggagctcgcqgtaacagcgtgcaaccccgagagtaa 438
Db 152680 CCTTTAAACAGGACCCTCCTACGCTGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAA 152739
QY 439 tgatccacggccttaatgtagagtccttggttcaacctttctctgtctcccgatcggtctga 498
Db 152740 TGATCACCAGCTTTAATGTAGGAGTCTCTGTCACCTTTTCTGCTCCCGATCGGTCTGA 152799
QY 499 tctgctcattgccactatctcagcagtggtgaaacagacagcgtcttcgtctgtatct 558
Db 152800 TCCTGCTCATTTGCCATATCTTCAGCAGTGTGTGAACAAGACAGCTCTTCTGCTGTATCTCT 152859
```

```
QY 559 cgcgcctgttgagtcctctgtgcagctggaaaattctactgcccgcgctcaacttgcgttg 618
Db 152860 CCCCCGTTCGAGTCCCTGTGCACTTGGAAATCTACTGCCGCGGTCAACTTGCCTGTGG 152919
QY 619 aggagatcggaactactacatacaaacaccttgcctctgcttgcgtggtgagcagagatggag 678
Db 152920 AGGAGATCGGATACTACATCACAAACCTTGTGCTGTGTTGTTGTCACGAGATGGAC 152979
QY 679 acgcctcgcgcctgtgatggaggatgtccctgtcacccgggtttggaataa----- 729
Db 152980 ACGCCTTGCCCTGTGATGGAGGATGTGCCCTGTACACCGGTTTGGAAATAAGAGTTAGAA 153039
QY 730 -----ag 731
Db 153040 TGATATAGCTACTTTACAAACCTTAACCCCTTACTTAATTAGGTAAATATTAATTTTTCAG 153099
QY 732 ttcatcttctgctgcgcttagcatatacaccgagctctccacacgaccttaaacagctcta 791
Db 153100 TTTCATCTTCTGCTGCGGTAGCATACACGGAGCTCTCCACGACCACCTTAAACAGCTTA 153159
QY 792 cgttggtccgcaactacgtattctgtgcgtggaatctgcataattttgtgttcgct 851
Db 153160 CGTTGTTCCGCAAGCTAGCTGTTCTGTGCGCTGGAACTTGGCATTAATTTTGTGTTGCTT 153219
QY 852 ggcgtgtgctatctcttaactcaacggtgggaatcactatgtcacctttgtacgcttac 911
Db 153220 GCGGTGTGCTATCTTCTTAATCTCAACGGTGGGAATCACTATGTACCTTTGTACGCTTAC 153279
QY 912 aaccacaacgtatggtgctactgaactaaacaggaataatcccgctgagggagagcgcgcc 971
Db 153280 AACCACAACGTAGTGTCTACTGAACCTAAACAAGAAATCCCGCTGAGGGAGAGCGCGCC 153339
QY 972 ttgcaagtggacaatacaataaaccgaagTaaacgctgcccagtaaacacgacgagagagt 1031
Db 153340 TTGCAAGTGGACAATCAATAATAACCCAAAGTAAACGGCTGCCAGTAAACAGCAGGAGAGT 153399
QY 1032 tgggtgacatgctgcagaaactctctgaagctcaacgcccgggtactgtgtgagtgcgagc 1091
Db 153400 TGGGTGACATGCTGTGAGAACTCTCTGAAAGTCAAGCCGGGTACTGTGTGAGTGGCGAC 153459
QY 1092 ttctgacgttaacgacgaaagcagcgcatctcaatcatagattggttaagctta 1151
Db 153460 TTTCGTGACGCTTAACGACGAAAGCAGCGGCATCTCAACATCATAGCATTTGATGGTCAAGTA 153519
QY 1152 cagtgctgtgataactaaatccgaacgttaaagctgcttcaggtggtgaggaacgcaaat 1211
Db 153520 CAGTGTGTGTAGTAAGTAATCCGAACGTAAAGCTGCTTCAGGTGGTGAGGACGCAAT 153579
QY 1212 ggagatgtcccggtggagctgcgcgacgatgtatgtctcaatgtgcgcgcactttggag 1271
Db 153580 GGAGATGTCCCGTGGAGCTGCGCAGCATGTATGTCTCAATGTGCGCGCGCACTTTGGAG 153639
QY 1272 gaggctccgagcactgctcgcgagtttgcaacgaggaattctgcctacgacgcctt 1331
Db 153640 GAGGTCTCCGAGCAGCTGCTGCGCGAGTTTGCACGAGGATTTGCCCTACGACCGCTT 153699
QY 1332 atacgaaatatacactgccataatgacgtttcaagcgacagaatttttcgcgagagagagctg 1391
Db 153700 ATACGAAATATCACATGCCATATATGAGTTCAAGCGACAGAAATTTTCGCGGAGAGAGCTG 153759
QY 1392 ccgcgcggtgatctatgtggggccatccatgggatgtcaactcgaactgtgaggtatccgcc 1451
Db 153760 CCGCGCGGTATCTATGTGGGCCATCCATGGGATGTCACTCGAATCTGGAGGTATATCCGCC 153819
QY 1452 ttgtgcgagatatagcttataaaggcagcctgagcgagtcgctgctgctccctc 1511
Db 153820 TTTGTGCCGAGATATAGCTTATTAAGGCAGCCTTGCCCCGATGCGCTGCTGCTCTCTC 153879
QY 1512 aagataaacgtgtcttcagcatagattggcgtgtgatacaatgccattccctgctttgct 1571
Db 153880 AAGTATAACGTGCTTTCAGCATAGGATTGGCGTTGATCAATGCCATTCCTCGCTTTGGT 153939
QY 1572 ttcatgtggcgcacaaatataccagcagcctgtgatacacagcttctgtgtggagagtgagat 1631
```



```

SEYERMDINDIYVPSPKYKTLIEVVLPSGNNNTTASGKNTPSDTQNDIQNDGPSDT
PNTSPSIDPTDPTDDEWNTLKDEFTSOYLQSEOPNDVPNDYSGNSNTINI
TTTSRHNVKEPFIMSJDRDLTYGEEYNDMSNNSGIYPSSSNDSLSGTPKVPYSGI
DLINSLSGNQPIDYDEILKRKENELFGTNHVYKQTSIHSVAKLTNSDPJHQLLELF
HKLDRHRDMEKLNDRNEAKLEENETHNSNTHIHTSDSNKLTNDVSIQIDMDID
GPKPKEFTNMOTILEDKYNEPYDVQDDIYDYVDNDHVDVSTVDSNAMDIPSKVQIEM
DYNTKLVKEKYPADWDI"
38960. 38965
/ gene="PFC0005w, MAL3p8.1"
/ note="Corrected splice donor for PFC0005c"
42416. 43015
/ note="R-FA3"
complement(43045..44255)
/ gene="PFC0010c, MAL3p8.2"
complement(join(43045..43971,44202..44255))
/ gene="PFC0010c, MAL3p8.2"
/ note="PFC0010c (MAL3p8.2), rfin (3D7-rift3-1), len: 326
aa; Similarity to other P.falci parum rifins. Member of a
multi-gene family; predicted using hexExon"
/ codon_start=1
/ protein_id="CAB39116.1"
/ db_xref="GI:4494021"
/ translation="MKIHYINILFELPLNLIYNQNHYYITRTPKATRTLCCELY
APSNYSDSPENKRVNQFVDRITQRFHEYDERMKTTRQCKEQCKEIKIILKDKLE
QMAQQLITLDPNITTEDIPICIEKSLADKTEKFLNCGKTMGVPAGWGLVSLGY
VGMWYVOTLALPGIEAGVAGIEELGFGPLRLIKVSOIQSIOPINPAVYAEFTYF
NFLKQVENTKCVKHEHTCYSTYISKENALSORASEIAKNAADMAKLAEREVLDKAAP
VTSGLNTALIASIVAILVILVLIYFLRYRKNKKKKLQYIKLLEE"
complement(46448..47828)
/ gene="PFC0015c"
complement(46448..47828)
/ gene="PFC0015c"
/ note="PFC0015c var c pseudogene"
/ codon_start=1
/ pseudo
complement(48956..49949)
/ gene="PFC0025c, MAL3p8.4"
complement(join(48956..49777,49881..49949))
/ gene="PFC0025c, MAL3p8.4"
/ note="PFC0025c (MAL3p8.4), stevor (3D7-stevorT3-1), len:
296 aa; Similarity to other P.falci parum stevors. Member
of a multi-gene family; predicted using hexExon"
/ codon_start=1
/ protein_id="CAB39117.1"
/ db_xref="GI:4494022"
/ translation="MKMYLKMFLTLNLTLPLPHYENLNNDYNVSEFIONTKRTT
IKSRLAQIQHNYHNDPELKEIILDKMNEAIKKYQOHDHPYQLKEVVEKNKTKY
SGNDAEPILTELEKLELTYEIEFGNEDMLKSGMNPVDEKSSFTFECTDINGVKLAK
TKGRDKYLLKLGKTRGTYFCGLSGVILTYIGTAAKGAATSTAGACKNCIYAI
FHLNHNKSSSALTALGTGVGCAADYASGATSASATFLPCGIAALVLLTLTLLIYI
WLYRRKRKRWKHECKHLCK"
complement(52066..53260)
/ gene="PFC0030c, MAL3p8.5"
complement(join(52066..53076,53192..53260))
/ gene="PFC0030c, MAL3p8.5"
/ note="PFC0030c (MAL3p8.5), rfin (3D7-rift3-2), len: 359
aa; Similarity to other P.falci parum rifins. Member of a
multi-gene family; predicted using hexExon"
/ codon_start=1
/ protein_id="CAB39118.1"
/ db_xref="GI:4494023"
/ db_xref="SPTREMBL:O97326"
/ translation="MKLIYVSNLILFALALNLTYSYAHNKNKPSITPHHRTTTSRL
LSECTESIYSDSEINSVKEILFERQTSRREEEYEMITQRQYKQRODKIOKIL
HKMKNEKIAEKGLRCGLGCVAGSGLGFAVAINWKPAALDAALAKATEG
TANISAAGVKAGEATGKVLVLSGLKMGISTLGGKNLESFYATTYSKKVASTAQAYVE
QHFEKCEFGYKGEPTGPDSPRDIHFQCSVMQQTQSVYKGTGHSYSPDKVTRVETM
YSKAGPANAFAEFYKATETATIKAAEKTITETASTOLYSAIGYSILAILIIVLIMI
VLIILYRKRKKKKKQYTKLNE"
58103. 59390
/ gene="PFC0035w, MAL3p8.6"
join(58103..58174,58338..59390)

```

```

/ gene="PFC0035w, MAL3p8.6"
/ note="PFC0035w (MAL3p8.6), rfin (3D7-rift3-3), len: 375
aa; Similarity to other P.falci parum rifins. Member of a
multi-gene family; predicted using hexExon"
/ codon_start=1
/ protein_id="CAB39119.1"
/ db_xref="GI:4494024"
/ db_xref="SPTREMBL:O97327"
/ translation="MRITKMKVHYINILLLALSINILANTHOKPSSTPRHIOITRELL
CEDIEHMYNDNDPENKRVNQFHDRTTQRFQYDERQKQVQKDCDKIEKILIL
KDKIEKELNEKFPATLOTDIHSDAIPCTICKTSMADKTEKLTGKLVGGIVAPSSGV
LAGIAEGALYAWKNALQATIAEALKAAADILVGGIEAGKVVIGGLDALIEKLG1
GSWEPTFTEGICINVKSLASIIYERKQTLGATKSTLDKATCEQISGTQHQHDTY
GAPGTPITETVLNGIVECTKEAADVAABARESATNAIKVQETRLLEAGFNSSISSIN
ASIAIVVILIMVIIIFILRYRRKKKKLQYIKLLEE"
61248. 62459
/ gene="PFC0040w, MAL3p8.7"
join(61248..61301,61548..62459)
/ gene="PFC0040w, MAL3p8.7"
/ note="PFC0040w (MAL3p8.7), rfin (3D7-rift3-4), len: 322
aa; Similarity to other P.falci parum rifins. Member of a
multi-gene family; predicted signal and transmembrane
sequence; predicted using hexExon"
/ codon_start=1
/ protein_id="CAB39120.1"
/ db_xref="GI:4494025"
/ db_xref="SPTREMBL:O97328"
/ translation="MKIHYINILFALPLNVLNQRNHNSTYHTSNTKTIKSHRSL
Query Match 2.3%; Score 45; DB 33; Length 108908;
Best Local Similarity 55.4%; Pred. No. 0.29;
Matches 87; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 1814 tcaccttcacgaacacacacacacagcgaatttagcacctcaaatcgcagtagct 1873
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101106 TCAAAATAAAGAACGAAACAAATAATAATAATAATAATAATAATAATAATA 101165
QY 1874 ttgttcagtagctccttagcttaactcgtatttatttctgcgcgtgttcgcagctcaaa 1933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101166 TTATATTATACANCGTTTITTTCTTATCTATAGGTAGCGCAAGACAGATT 101225
QY 1934 ataaatcaaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1970
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101226 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 101262
RESULT 9
AC013583 173184 bp DNA HTG 16-MAR-2000
LOCUS Homo sapiens clone RP11-5C23, WORKING DRAFT SEQUENCE, 18 unordered
DEFINITION pieces.
ACCESSION AC013583
VERSION AC013583.4 GI:7248956
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173184)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,M.,
AUTHORS Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
TITLE Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
JOURNAL Cooke,P., DeArnell,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
REFERENCE Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
AUTHORS Galagan,J., Gardyana,S., Grant,C., Hagos,B., Heaford,A., Horton,L.,
TITLE Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
REFERENCE Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
AUTHORS McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
TITLE Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
1 (bases 1 to 173184)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
AUTHORS Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
TITLE Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
JOURNAL Cooke,P., DeArnell,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
REFERENCE Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
AUTHORS Galagan,J., Gardyana,S., Grant,C., Hagos,B., Heaford,A., Horton,L.,
TITLE Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
REFERENCE Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
AUTHORS McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
TITLE Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
1 (bases 1 to 173184)

```

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6970296.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2847
Center clone name: 5_C_23
----- Summary Statistics
Sequencing vector: M13: M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160008 bases at least Q40
Consensus quality: 166750 bases at least Q30
Consensus quality: 169341 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 171484; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1031: contig of 1031 bp in length
* 1032 1131: gap of 100 bp
* 1132 3436: contig of 2305 bp in length
* 3437 3536: gap of 100 bp
* 3537 5182: contig of 1646 bp in length
* 5183 5282: gap of 100 bp
* 5283 8761: contig of 3479 bp in length
* 8762 8861: gap of 100 bp
* 8862 14445: contig of 5584 bp in length
* 14446 14545: gap of 100 bp
* 14546 20653: contig of 6108 bp in length
* 20654 20753: gap of 100 bp
* 20754 26231: contig of 5478 bp in length
* 26232 26331: gap of 100 bp
* 26332 34259: contig of 7928 bp in length
* 34260 34359: gap of 100 bp
* 34360 40262: contig of 5903 bp in length
* 40263 40362: gap of 100 bp
* 40363 46964: contig of 6602 bp in length
* 46965 47064: gap of 100 bp
* 47065 56070: contig of 9006 bp in length
* 56071 56170: gap of 100 bp
* 56171 71111: contig of 14941 bp in length
* 71112 71211: gap of 100 bp
* 71212 80835: contig of 9624 bp in length
* 80836 80935: gap of 100 bp
* 80936 95321: contig of 14386 bp in length
* 95322 95421: gap of 100 bp
* 95422 108414: contig of 12993 bp in length
* 108415 108514: gap of 100 bp
* 108515 122869: contig of 14355 bp in length
* 122870 122969: gap of 100 bp
* 122970 137578: contig of 14609 bp in length
* 137579 137678: gap of 100 bp
*

FEATURES
source

misc_feature 1..1031

misc_feature 1132..3436

misc_feature 3537..5182

misc_feature 5283..8761

misc_feature 8862..14445

misc_feature 14546..20653

misc_feature 20754..26231

misc_feature 26332..34259

misc_feature 34360..40262

misc_feature 40363..46964

misc_feature 47065..56070

misc_feature 56171..71111

misc_feature 71212..80835

misc_feature 80936..95321

misc_feature 95422..108414

misc_feature 108515..122869

misc_feature 122970..137578

misc_feature 137679..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

misc_feature 173185..173184

Db 6153 ATT 6155

RESULT 10

ALI36310/c

LOCUS

DEFINITION

ALI36310 237302 bp DNA HTG 20-APR-2000

HOMO sapiens chromosome 6 clone RP3-415N12, *** SEQUENCING IN

PROGRESS ***, in unordered pieces.

ACCESSION

ALI36310

VERSION

ALI36310.5

KEYWORDS

HTG; HTGS-PHASE1.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Burton.J.

Direct Submission

Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Apr 21, 2000 this sequence version replaced gi:7378514.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: dJ415N12

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 193992 bases at least Q40

Consensus quality: 210387 bases at least Q30

Consensus quality: 220937 bases at least Q20

Insert size: 230202; sum-of-contigs

Insert size: 137822; 12.5% error; agarose-fp

Quality coverage: 2.65x in Q20 bases; sum-of-contigs Quality

coverage: 4.43x in Q20 bases; agarose-fp

* NOTE: This is a 'unfinished' sequence. It currently * consists of 72 contigs. The true order of the pieces is * not known and their order in this sequence record is * arbitrary. Where the contigs adjacent to the vector can * be identified, they are labelled with 'clone end' in the * feature table. Some order and orientation information * can tentatively be deduced from paired sequencing reads * which have been identified to span the gap between two * contigs. These are labelled as part of the same * 'fragment_chain', and the order and relative orientation * of the pieces within a fragment chain is reflected in * this file. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as * soon as it is available and the accession number will be * preserved.

* 1 3938 contig of 3938 bp in length; fragment_chain 1 *
4039 7596 contig of 3558 bp in length; fragment_chain 1 *
7697 25193 contig of 17497 bp in length; fragment_chain 1 *
25294 48416 contig of 23123 bp in length; fragment_chain 1 *
48517 79454 contig of 30938 bp in length; fragment_chain 1 *
79555 88282 contig of 8728 bp in length; fragment_chain 2 *
88383 92814 contig of 4432 bp in length; fragment_chain 2 *
92915 94176 contig of 1262 bp in length; fragment_chain 2 *
94277 96619 contig of 2343 bp in length; fragment_chain 2 *
96720 98893 contig of 2174 bp in length; fragment_chain 3 *
98994 100558 contig of 1565 bp in length; fragment_chain 3 *
100659 102748 contig of 2090 bp in length; fragment_chain 3 *
102849 104488 contig of 1640 bp in length; fragment_chain 4 *
104589 105793 contig of 1205 bp in length; fragment_chain 4 *
105894 109086 contig of 3193 bp in length; fragment_chain 4 *
109187 110267 contig of 1081 bp in length; fragment_chain 5 *
110368 111393 contig of 1026 bp in length; fragment_chain 5 *
111494 112651 contig of 1158 bp in length; fragment_chain 5 *

112752 114034 contig of 1283 bp in length; fragment_chain 6 *
114135 115953 contig of 1819 bp in length; fragment_chain 6 *
116054 117793 contig of 1740 bp in length; fragment_chain 7 *
117894 119191 contig of 1298 bp in length; fragment_chain 7 *
119292 120448 contig of 1157 bp in length; fragment_chain 8 *
120549 122011 contig of 1463 bp in length; fragment_chain 8 *
122112 123814 contig of 1703 bp in length; fragment_chain 9 *
123915 125651 contig of 1737 bp in length; fragment_chain 9 *
125752 126854 contig of 1103 bp in length; fragment_chain 10 *
128955 129043 contig of 2089 bp in length; fragment_chain 10 *
129144 130148 contig of 1005 bp in length; fragment_chain 11 *
130249 131530 contig of 1282 bp in length; fragment_chain 11 *
131631 134360 contig of 2730 bp in length; fragment_chain 12 *
134461 139505 contig of 5045 bp in length; fragment_chain 12 *
134606 141471 contig of 1866 bp in length; fragment_chain 13 *
141572 143919 contig of 2348 bp in length; fragment_chain 13 *
144020 145706 contig of 1687 bp in length; fragment_chain 14 *
145807 149597 contig of 3791 bp in length; fragment_chain 14 *
149698 152091 contig of 2394 bp in length; fragment_chain 15 *
152192 153223 contig of 1032 bp in length; fragment_chain 15 *
153324 154577 contig of 1254 bp in length; fragment_chain 15 *
* 154678 156040 contig of 1363 bp in length
* 156141 157869 contig of 1329 bp in length
* 157770 158989 contig of 1220 bp in length
* 159090 169641 contig of 10552 bp in length
* 169742 171499 contig of 1758 bp in length
* 171600 172750 contig of 1151 bp in length
* 172851 174805 contig of 1955 bp in length
* 174906 177316 contig of 2411 bp in length
* 177417 178643 contig of 1227 bp in length
* 178744 182008 contig of 3265 bp in length
* 182109 184328 contig of 2220 bp in length
* 184429 186132 contig of 1704 bp in length
* 186233 190452 contig of 4220 bp in length
* 190553 194819 contig of 4267 bp in length
* 194920 197185 contig of 2266 bp in length
* 197286 198667 contig of 1382 bp in length
* 198768 201386 contig of 2619 bp in length
* 201487 202629 contig of 1143 bp in length
* 202730 203933 contig of 1204 bp in length
* 204034 206262 contig of 2229 bp in length
* 203633 209561 contig of 3199 bp in length
* 209662 212772 contig of 3111 bp in length
* 212873 214378 contig of 1506 bp in length
* 214479 218263 contig of 3785 bp in length
* 218364 221064 contig of 2701 bp in length
* 221165 222898 contig of 1734 bp in length
* 222999 224757 contig of 1759 bp in length
* 224858 226013 contig of 1156 bp in length
* 226114 228699 contig of 2586 bp in length
* 228800 230694 contig of 1895 bp in length
* 230795 233315 contig of 2521 bp in length
* 233416 234443 contig of 1028 bp in length
* 234544 237302 contig of 2759 bp in length.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source

Location/Qualifiers

1. 237302

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP3-415N12"

/clone_lib="RPCI-3"

1. 3938

/note="assembly_fragment:01023"

fragment_chain:1"

4039..7596

/note="assembly_fragment:01368"

fragment_chain:1"

7697..25193

/note="assembly_fragment:00456"

misc_feature

misc_feature

misc_feature


```

/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="91F10-91F11"
/clone="DS02393 (D521)"
/clone_lib="P1 library, partial Sau3A in pAd10sacBII"
BASE COUNT 13212 a 10748 c 10575 g 13740 t 2961 others
ORIGIN

Query Match      2.2%; Score 44.2; DB 41; Length 51236;
Best Local Similarity 61.9%; Pred. No. 0.43;
Matches 70; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1858 caaagattcgatagctttttgtcatagctcttagcttaactcgctatttatttttcgtacg 1917
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 50240 CAACGCTTCAAAAAATACCTGTGTATAGTTTAAGATTTTAACTTATTGTTAGTCTCTATA 50181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1918 gttgtcgagctcaaaaatacaataaagctcaaaaaaataaaaaa 1970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 50180 CAAGAAGATTCAATAATAAACCAAGTAATAATAAAAAA 50128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AC004336/c
LOCUS
DEFINITION Drosophila melanogaster DNA sequence (P1 DS03544 (D234)), complete
sequence.
AC004336
VERSION AC004336.1 GI:3492863
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in sac from P1 clone DS03544
(D234)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 61267)
AUTHORS Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2R, region 55A2-55B2
Unpublished (1997)
2 (bases 1 to 61267)
Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Aug 29, 1998 this sequence version replaced gi:3419813.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley Drosophila Genome project
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://fruitfly.berkeley.edu/sequence/) or send
email to bdgp@fruitfly.berkeley.edu.
Library location: 88-37
Location/Qualifiers
1..61267
/organism="Drosophila melanogaster"
FEATURES
source

```

```

/db_xref="taxon:7227"
/chromosome="2R"
/map="55A2-55B2"
/clone="P1 DS03544 (D234)"
/note="DS03544 (D234) is a bridge overlapping DS08860
(D181) and extending to pl end."
BASE COUNT 17069 a 12971 c 13472 g 17755 t
ORIGIN

Query Match      2.2%; Score 44.2; DB 34; Length 61267;
Best Local Similarity 61.9%; Pred. No. 0.44;
Matches 70; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1858 caaagattcgatagctttttgtcatagctcttagcttaactcgctatttatttttcgtacg 1917
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20209 CAACGCTTCAAAAAATACCTGTGTATAGTTTAAGATTTTAACTTATTGTTAGTCTCTATA 20150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1918 gttgtcgagctcaaaaatacaataaagctcaaaaaaataaaaaa 1970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20149 CAAGAAGATTCAATAATAAACCAAGTAATAATAAAAAA 20097
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: September 4, 2000, 00:00:24
Job time: 19444 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:36:20 ; Search time 11551.9 Seconds
(without alignments)
528.216 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggttaattaccacagtttg.....taaatcagattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: em_fun:*
- 17: em_hum1:*
- 18: em_hum2:*
- 19: em_in:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_sy:*
- 29: em_un:*
- 30: em_vi:*
- 31: gb_htg1:*
- 32: gb_htg2:*
- 33: gb_in1:*
- 34: gb_in2:*
- 35: em_bal:*
- 36: em_ba2:*
- 37: em_hum3:*
- 38: em_hum4:*
- 39: gb_pr4:*
- 40: gb_htg3:*
- 41: gb_htg4:*
- 42: gb_htg5:*
- 43: gb_htg6:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_pl3:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	1008.4	29.5	95968	33	CEY47D3B	AL031635 Caenorhab
C 2	1008.4	29.5	197735	31	CEY66A7	AL022282 Caenorhab
C 3	1008.4	29.5	337565	31	CEY47D3	Z98865 Caenorhabdi
C 4	604.8	17.7	36780	31	CEH10N23	Z94157 Caenorhabdi
5	535	15.6	296699	31	CEY48A6	Z92854 Caenorhabdi
6	424.2	12.4	36780	31	CEH10N23	Z94157 Caenorhabdi
C 7	71.8	2.1	7218	5	I66494	I66494 Sequence 14
8	57.4	1.7	3816	12	CGU09103	U09103 Cricetus
9	56.8	1.7	1209	12	AB017337	AB017337 Mus muscu
10	55.6	1.6	1767	12	CGU22819	U22819 Cricetus
11	55.6	1.6	1914	12	CGU12329	U12329 Cricetus
12	55.6	1.6	4425	12	CGU22818	U22818 Cricetus
13	55.6	1.6	4987	12	GUL2330	U12330 Cricetus
14	52.6	1.5	3115	12	RATADD1A	L16995 Rat addl mR
15	51	1.5	4154	5	AR069298	AR069298 Sequence
16	51	1.5	4154	5	I22310	I22310 Sequence 37
17	51	1.5	4154	50	U00968	U00968 Human SREBP
18	49.4	1.4	3768	33	DMU38238	U38238 Drosophila
19	49.4	1.4	44517	42	AC018013	AC018013 Drosophi
C 20	49.4	1.4	273521	34	AE003515	AE003515 Drosophi
C 21	47.8	1.4	92677	51	AC009382	AC009382 Drosophi
22	47.6	1.4	4249	9	HSU02031	U02031 Human stero
23	47.6	1.4	5197	5	AR069307	AR069307 Sequence
24	47.6	1.4	5197	5	I22319	I22319 Sequence 53

```

25 46.8 1.4 142947 55 AC010488 Homo sapi
26 44.6 1.3 127485 50 HUAC02039 Homo sapi
27 43.8 1.3 433 3 AF102873 Sus scrofa
28 43.8 1.3 976 7 SOYNOD75 Soybean ear
29 43.8 1.3 3060 5 I08177 Sequence 1
30 43.8 1.3 3072 7 GMEOD2A
31 43.8 1.3 4584 7 GMEOD2B
32 43.8 1.3 184452 40 AL138924 Homo sapi
33 43.6 1.3 156107 78 AC016898 Homo sapi
34 43.6 1.3 176929 39 AC004887 Homo sapi
35 43 1.3 45993 1 AB012957 Vibrio chol
36 42.6 1.2 29605 7 AB020747 Arabidopsis
37 42.6 1.2 191375 55 AC023167 Mus muscu
38 42 1.2 40351 11 AC005620 Homo sapi
39 41.8 1.2 181671 31 AP001024 Homo sapi
40 41.4 1.2 12668 34 AF079445 Dictyoste
41 41.4 1.2 14363 1 VCGTNRGN Vibrio chol
42 41.4 1.2 46721 1 AB012956 Vibrio chol
43 41.4 1.2 90765 73 AC010483 Homo sapi
44 41.4 1.2 98491 12 AF132039 Mus muscu
45 41.4 1.2 168239 60 AC021367 Homo sapi

```

ALIGNMENTS

```

RESULT 1
CEY47D3B/c
LOCUS CEY47D3B 95968 bp DNA INV 14-DEC-1999
DEFINITION Caenorhabditis elegans cosmid Y47D3B, complete sequence.
ACCESSION AL031635.2 Z98865
VERSION AL031635.1 GI:3646936
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1
AUTHORS Rhabditia; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
TITLE Rhabditia; Rhabditidae; Rhabditidae; Pelodera; Caenorhabditis.
JOURNAL 1 (bases 1 to 95968)
MEDLINE Genome sequence of the nematode C. elegans: a platform for
REMARK investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
The C. elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1;283(5398):35 and 1999 Mar 26;283(5410):2103]]
2 (bases 1 to 95968)
Mathews,L.
Direct Submission
Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence&object=Y47D3B
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y47D3B.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.

```

```

FEATURES
source
1. 95968
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y47D3B"
/clone="Y47D3B.1"
/gene="Y47D3B.1"
join(7188..7236,7287..7436,7488..7605,7654..7713,
8884..9277,10132..10318,11510..11678,11731..11860,
12379..12481,12657..12736)
/gene="Y47D3B.1"
/codon_start=1
/protein_id="CAA21039.1"
/db_xref="GI:3881005"
/db_xref="SPTREMBL:O9XX03"
/translation="MWAKNIGDTFLRKDINTROIILGFWDEQNVDCGRFAVIHF
LGDRIEHLPLCYFSIQANNRYLITKATVORHGRKASSDICSNAGHIAECDLDVHAL
KEFKITTPNVEDAMLTPELPREPHDHYVCMAPVIYTDWEILVTGIELWAMGA
TKIVPQASNATYRILOEYERKGLVIRTPKWPIMSDTNPGLVLSRGIEESHVN
CLFEVQWSDIVFSDIDDLPLDPSITSPGNDLQIKNFIAHPQAGSLFHEHDV
QFVPPNQGDSLTNFENFEFURSSKNMNCNWRMKTTRVNVASRDVSNMHTGIDR
EYVQTVRVPQRAHYHLRHSHTVPSPPTINMSPLADMLANKQOTRVEGTGMKDEH
LNKNTYELEDRCMGAINDEHMTKVSRCLTTHVCFSLRRLREVDCVAVASEYEFYR
SGKRIQAEVNCESPVPFPFTSGNHYFAP"
complement(12895..17349)
/gene="Y47D3B.2a"
complement(join(12895..12950,13560..13955,17307..17349))
/gene="Y47D3B.2a"
note="cDNA EST yk265b3.5 comes from this gene; cDNA EST
EMBL:Cl0452 comes from this gene; cDNA EST EMBL:Cl2093
comes from this gene; cDNA EST EMBL:D70235 comes from this
gene; cDNA EST yk552e6.3 comes from this gene"
/codon_start=1
/protein_id="CAA21040.1"
/db_xref="GI:3881006"
/db_xref="SPTREMBL:O9XX02"
/translation="MRNSLFTTLFGLAALVWLNAQYTSLEDEKRGGRARMLHHR
GGARAFSAVDGDDYKRGARAFYDEKRGARAFTEMKRGARVFGDEKRGARA
FMMDKRGGRGAFGDMKRGARAFVENSKRDEWVIRPDEDDRLKRGGRSFPVKFG
RLDD"
complement(13549..17349)
/gene="Y47D3B.2b"
complement(join(13549..13955,17307..17349))
/gene="Y47D3B.2b"
/codon_start=1
/protein_id="CAB54383.1"
/db_xref="GI:5824792"
/translation="MRNSLFTTLFGLAALVWLNAQYTSLEDEKRGGRARMLHHR
GGARAFSAVDGDDYKRGARAFYDEKRGARAFTEMKRGARVFGDEKRGARA
FMMDKRGGRGAFGDMKRGARAFVENSKRDEWVIRPDEDDRLKRGGRSFPVKFG
20481..30076"
/gene="Y47D3B.3"
join(20481..21742..22022,22112..22190,23753..23852,
24487..24611,24655..24756,25563..25693,25796..25891,
27970..28341,29197..29332,29684..29742,29801..30076)
/gene="Y47D3B.3"
note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21038.1"
/db_xref="GI:3881004"
/db_xref="SPTREMBL:O9XX04"
/translation="MKNKILKVRSLGFIFGONFLISKYLRFCISLLDPQLDLKLQT
RMVPVPFAIIVSFNFSLIFNHTYISHVATLSLSFSFFFRHIFPHLPOQSTSF
QEVFLYSVLLVIVRSGIAVDYDIEILPENCDCLDYCKACKVNGKVDWYKYS
DGKRSDATFTGCGQGVADVAGQNTIEDLMQVCTDAOKLHYETIEPCFENGTKIR
VGKEVQDGSFQWLCLSTGRWITGCVYFNETHDILLRVDEHAYNGLIIEHVCSKROEYP

```

The true left end of clone H10N23 is at 61856 in this sequence. The true right end of clone T28D6 is at 104 in this sequence. The true right end of clone Y47D3 is at 95968 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z81134. The end of this sequence (95863..95968) overlaps with the start of sequence AL032622.

AIVOYKTOVRKDVVDVHPTNKGYNRNPPPELQKLIQDFIDINTIFLSCFCVSCNTTETS
TTTTPIPHSTSEAPLCTRDYHPCSDVCRSGICTHLLNCSRDYHCPACEVCRSGK
CIGGEQKCKTEKCEAGCWCMEGCKVENKDCIACVERAICKVGLEOIFEXYVS
LNFNEYWGEHAQOOSNFVEPCRTIPAMNVSTFIKLLIFIFIFPKFKETKFGREENSP
DYDNEDAKRDELIFVKNYUKMQGRAPAPVPVKKHNVEESTWCLSPPKNNVRKPRS
WLMKAGIEVYSAYEDDRENSLFPGNSAIQV*
/gene="Y47D3B.4"
/join(30854..32478
/gene="Y47D3B.4"
/join(30854..30889,31202..31345,31929..32127,32258..32478)
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21037.1"
/db_xref="GI:3881003"
/db_xref="SPTREMBL:O9XX05"
/translation="MPMCLHIAQLLEDISGRLLLEWQYLFQADSVLSIHYPLTNQ
THKVLQYKHTNLKVIPLQSDNGPNSPISERSKYLENPKQRKRELLIPLYNOCFYNY
IPIHRYVLIDIDIVLPIQFODNYGDLDRDFEAKSFGYRLSSISSNRNFKFSPNFTIF
PRHYMISNRKSRKSTKGEYKGSFSSRYRELFCLIFA"
34913..48475
/gene="Y47D3B.5"
/join(34913..35016,38489..38669,39530..39640,39697..39824,
39884..40076,40805..40983,41034..41172,42083..42283,
42356..42498,42550..42774,43192..43259,43308..43485,
43538..43666,44148..44300,44351..44419,44474..44658,
44702..44801,45577..45675,45727..46265,47093..47246,
47606..47707,48365..48475)
/gene="Y47D3B.5"
/note="cDNA EST yk449el0.5 comes from this gene; cDNA EST
EMBL:D35319 comes from this gene; cDNA EST EMBL:C12322
comes from this gene; cDNA EST EMBL:D32633 comes from this
gene"
/codon_start=1
/protein_id="CAA21041.1"
/db_xref="GI:3881007"
/db_xref="SPTREMBL:O9XX01"
/translation="MPNPSPTPLQASPTTTLLADGYTDDALPSKYDNFRASRPSRES
GGFSRAGSLVEGTGIVGPPMSPISFLGCKEYHYDKYKLDRIASVLLVLYSLFCAW
VYLFEDYEREVKLERIDLMRLNDTQRISSWVFRQGLANFEVFLDYELKLVH
VRLPECLDNDWYGALFYVGTFTTIGYNIYPRTAGRAASVVAIVGLPLVLALSK
CGKMTDSEKQWQHRIQIEKAKMTKMRKRGKILKSGEIVAEANTGAPAEKPEKE
VESRTIPLWALLICVYVAGSSFLKMETWRTFTTSLYFFCISLITGLIGDIVPK
PHMFIYMFVLIVGLSVSMFISVWQIKIEWLCRMTEQTKERYMKKDGILVDEPIKR
IPANDVSLFVAPKINSGEONHMKDITIKKEDRLLESKEIQTDPFVKGNATOVEKYEQ
SMACDPSNHLKPAHQGTQWSPVDETGEGPRNVELKLQRDDATISRSNDPRSI
SDATSLPMSMSSPVAKRKNRNTCEVACQCDLPMGNTAEFRADGTAQIDLQAFQIDE
IATLANLOTQVRPPIVEMATSAFMEQSDPLNSEALDLPRLSQADINVICEAMSSSF
FKNSQALQIADIGYKDYETTFSDKSQLTSLTPVPTRNISIAISPVYSLCEDLKA
SKKSKSRSTSSSPIPNLSHEEETQPLRNQVHMATDPTCSQLVHRSTVTSVPGRE
DAPTEMPOKMDRMVSPVSTAIRONTSPVEROTPDFDORSIOTSISEWFGKLLKKKD
ESQOTLCPNDVYVCKKKEGKKKASRDENSQASSLGSVDFTSSNPSSTSDDSKNNKK
IKOTLIAPNVSASVSTQSPPPGFDRLSEPHSTRASSESTSGMGTSLFEDHTRQEV
IVTDDSYLKIARLDEIRNNKTQFLFPCAAPLSKEVPEPFKSDRPSERSHYFQGG
RASLGRGRKQDMRYGESQGTQSMQDAELLKELVLSSEKRSISPKRPNCRKRGVLSR
HPSLPTGISRGKGVSVIQHEKGIHNPATNRQSPIKIIRQYSLDHDVI"
51392..52342
/gene="Y47D3B.6"
/join(51392..51449,51544..51641,51689..52342)
/gene="Y47D3B.6"
/note="predicted using GeneFinder; cDNA EST yk506e5.3
comes from this gene"
/codon_start=1
/protein_id="CAA21043.2"
/db_xref="GI:5824793"
/translation="MFLRTICATLLAFTAAQVPVAPVLYSACSAAVPCAAVQASVC
IPTAAAPPMMAPPLPAPVAPAPLPVAPVFAAPVFAAPVFAAPVFAAPVFAAPV
IAAQAALCAAPACGLGAOCVAGACANPAVYSPVIGCOPALPAPVAPVACRLPIQ
ALPGSPCEPVECTGSGVSVGLCPPELLIOBCTVCVARTIYGVVPPVPIVPVAPV
VALGAPCAAPVAPVYASAPVFGAVCSAGVCCGAGVLEPLQAGVRRRL"
/complement(52583..566730)
/gene="Y47D3B.7"
complement(join(52583..52774,52830..53190,61122..61275,

61322..61549,61609..62005,62315..62843,62889..63082,
63606..63841,64186..64605,64659..64767,65539..65848,
66107..66239,66652..66730))
/gene="Y47D3B.7"
/note="predicted using GeneFinder; similar to
Helix-loop-helix DNA-binding domain; cDNA EST EMBL:D34815
comes from this gene; cDNA EST EMBL:D35004 comes from this
gene; cDNA EST EMBL:D68853 comes from this gene; cDNA EST
EMBL:D67853 comes from this gene; cDNA EST yk349h3.5 comes
from this gene"
Query Match 29.5%; Score 1008.4; DB 33; Length 95968;
Best Local Similarity 76.6%; Pred. No. 1.6e-280;
Matches 1508; Conservative 0; Mismatches 1; Indels 459; Gaps 4;
Qy 1304 attaccgacccggaaccaattcaatacactgaataactctcccgatcacccgctgaatc 1363
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 63089 ATTACGTACCGGAACCAATTCGAATACACTGAATACTCTCCCGATCACCGCTCGAATC 63030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1364 atctcctctccacttagaataagagaaaaacgatacagaatgagcaacaacgactcctat 1423
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 63029 ATCTCCTTCTCCACTAGAAATGAGAGAAAACGATCAGCAATGAGCAACAGCACTCTAT 62970
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1424 gaagaatggaaactagatggaatcttcgaaagttaccccttttttcgtagtctccatagcagt 1483
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62969 GAAGAATGGAACCTAGAGATGGATCTTCGAAAGTTTACCCTTTTTCGATGCTCTTAGCAGT 62910
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1484 tctgattttlaatccgattg----- 1503
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62909 TCTGATTTTTAATCCGATTGGGTAAAGCTTCTATTCTGAAAAATAGCGTTAAATCCTT 62850
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1504 ----gattgctcctggaaagtcgatatctcctcaaaagccgctcgaagctccgattgc 1558
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62849 GTTTAGATTCTCCTCGCTGGAAGTGCGATATTTCTCAAAAGCCCTCCAGAGCTCCGATTGC 62790
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1559 ctcccccttcgagcatggaagagtgattgtagccccggatggaactagcactcgagcgt 1618
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62789 CTCCCGCTTCGAGCATGGAAGAGTGATTGATGACCCCGATGGAACCTAGCAGCTCGACGCT 62730
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1619 ttctggaaaggagatcatcaatagatgagatactgtctgggtgttcaacatcttaataatg 1678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62729 TTTCTGGAGAGGAGTATCATCAATATGAGTATGCTCTGGGTGTTTCAACATCTTAAATGAT 62670
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1679 catatatggtgttctcaaaactgctgataccatggtagccctgttcaagacttcatgtccgt 1738
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62669 CATATATGCTGTTCTCAACTGCTGATCCATGCTGACCTCTTCAAGACTTTCATGTCGT 62610
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1739 ttcattggcaacttttctgacgactcagagaaagcgagagccgaagtgaactctgaaa 1798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62609 TTCATGGCAGACTTTTGTGACGACTCGAGAGAGGCGGAGAGCCGAGTTGAACTCTGAAA 62550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1799 ttgaaagatcctcagaaaaattctcgagatgtcttgcaacgttgatcgatcgcttc 1858
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62549 TTTGAAAGATGCTCAGAGAAAGTTTCTCGAGTGTCTTGCACAGTTGGATCGATCGCTTCC 62490
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1859 atcacccgggggtgattcgggtgttttcgggtggagaaagcgttcgacatctttttaa 1918
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62489 ATCACCGGGGCTGATTCGGTGTGCTGGCTGGGAATCGCTTCGACATCTTTGAA 62430
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1919 ttgatttgatcgaggagatacatcgcaagaagcgaggtccaccacgaagcgtgtctc 1978
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62429 TTGGTTGTGGATCGGAGATACATCGCAAGAGGCGAGGTCACACCAAGCCCTGCTTC 62370
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1979 agtcgtttgtaggaagtcacgcgaagactgcagttctctatcatgaaattcat----- 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62369 AGTCGTTGTAGGAGTATCGCGAGACTGCGAGTCTCTATCATGAAATTCATCAGGTTTA 62310
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 2031 ----- 2030
Db 62309 CTGTCAAAATTCATAAATTTTGCATTTCAAAAATATATTTTACTAGAGTAATCTACAGTAC 62250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 2031 ----- 2030
Db 62249 TGCTACAGTACCCCTACAGTACTACTACAGTACCCTTGACCATATATCCCACTAAGTCCAAA 62190
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```
Qy 2031 ----- 2030
Db 62189 CTAATATCCCTCAAAGACAAAAAGTCAATTTTCCAAAACACAGTAATCCCTTCAGTA 62130
Qy 2031 ----- 2030
Db 62129 CTCCTACGGTACCATTAACAGTACTACTACAGTACCCCTGACCATATCCCCCACCATACTCC 62070
Qy 2031 ----- 2030
Db 62069 GAACATAATATCCCTCAAAGACAGACAAAAACATGATTTTATGTATAAGTAATGATTTTT 62010
Qy 2031 -cagctccatcaatggtatcactggaactcgaacacacctgaacacctgaacctcgcgcct 2089
Db 62009 CCAGCTCATCTAATGGGTATCAGCTGGAACCTCGAAGACACCTATGAACCATCCGCCCT 61950
Qy 2090 aacgggctcttcatgctccctctgtgcagtaaaccttgcgaagctgcgcgagacataaa 2149
Db 61949 AACGGGCTCTTCATGTCCTCTGTGCAGTAAACCTTGTCTGAAGCTGCCGGAGCATCAA 61890
Qy 2150 gacggaactccacgcgcgtcatggtcagatctacatcttctgcatccatccatccatgcgcg 2209
Db 61889 CGACGGACTTCACGCGCGCTCATGGCTCAGATCTACATTTCTGTCATCCATCCATGCCG 61830
Qy 2210 ttggctcttcgaacctactgcgaccattctctcgggatacttttacgaagagctcg 2269
Db 61829 TTTGGCTCTTCGAACCTACTCGCACCATTTCTCTCGGGATACTTTTCAGGAAGAGCTCG 61770
Qy 2270 aaggcactgpgctgagctcgggagcactcgcgtgtccattgttgatgattctccatcc 2329
Db 61769 AAGGCACGTGGCTGAGCTCGGAGCACTCGGTGTCCCATTTGTTATGGATCTTCATCC 61710
Qy 2330 agcgaacaagaagtctcatgctcagatgcgaagaaaggttgagcatggttgagctcgaagca 2389
Db 61709 AGCGACAAGAAAGTTCATGTCAGATGCGAAGAAAGTTGGAGCATGTTGTTGAGCTCGAAGCA 61650
Qy 2390 gaagcagttgagattgggtctttgtggaagatgagcaat----- 2430
Db 61649 GAAGCAGTTGAGATTGGTCTCTTTGTCGAGATGAGCAATCTAAGTATTTTAAATAAA 61590
Qy 2431 -----tatccccactgtcgaatc 2450
Db 61589 ATAAAAATAAAATAAAATAAACTTTTCATTTTCAGTATCCCCACTTGTCTCGAATC 61530
Qy 2451 caacaacgcgtgaagtgtacctactctccaaactgtacaggaactgtcgtggtgac 2510
Db 61529 CGAACACGCTGAAAGTGTACCTACTCTCCAAACTTGTACAGGAACCTTGTGCGTGGTGAC 61470
Qy 2511 gagatctttacaaaaatgtggaacgcatccttaaatgacaatgacctctcgtgatgaa 2570
Db 61469 GAGATCTTTACAAAAATGTGGAACGCATCTCTAAATGACAATGACCGTCTCGATGATGAA 61410
Qy 2571 gtgacgtggttgatglttccaagactttgttgacaaattccaacgagctgagctgcaatt 2630
Db 61409 GTAGACGTGGTGTGTTTCAAGACTTTTGTGTGACAATTTTCAACGACAGTGGCGTGCCTAT 61350
Qy 2631 ttgactaatgagaagatgagtccagca----- 2658
Db 61349 TTGACTAATGAGAGGATGAGTGCAGCGAGTTAGTGAACACGCAATTTTAAAGATTGAAATG 61290
Qy 2659 -----aatcgcgaacctggtatctctcgaacgagatgctgttgacatg 2704
Db 61289 TTATAAATTTCTAGAAATTCGGAACCTGGATCTCTCGAAGCGAGATGCTTGTGTCACATG 61230
Qy 2705 gtggacgacgtctctgacatgtggaatctattggaggagtgaaagaatgagctggcagc 2764
Db 61229 GTGGACGACGCTCTGACATGTGGAATCTATTGGAGGAGTAAACAAGATGAGCTGGCACG 61170
Qy 2765 gcaacactattcactgatcagaacctgtcccgccgaagattttgacaga 2812
Db 61169 GCAACACTATTACATGATCAGGAACCTGTCCGCCGAAGATTGTTGACAGA 61122
```

```
RESULT 2
CEY66A7/c
LOCUS CEY66A7 197735 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans chromosome III clone Y66A7, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION AL022282
VERSION AL022282.1 GI:3451532
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 197735)
Matthews, L.
Direct Submission
Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
On Aug 25, 1998 this sequence version replaced gi:3250755.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source            1..197735
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="III"
                        /clone="Y66A7"
BASE COUNT 62861 a 35824 c 35278 g 62783 t 989 others
ORIGIN
```

Query Match 29.5%; Score 1008.4; DB 31; Length 197735;
Best Local Similarity 76.6%; Pred. No. 1.9e-280;
Matches 1508; Conservative 0; Mismatches 1; Indels 459; Gaps 4;

```
Qy 1304 attaccgtaccggaaacaaattcaatcacactgaatactctcccgatcacccgtcgaatc 1363
Db 4043 ATTACAGTACC CGGAACCAATTCAATACACTGAA TACTCTGCCGATCACC CGTGAATC 3984
Qy 1364 atctctctccacctagaatgagaaaaacgatacgaatgagacaacgactcctat 1423
Db 3983 ATCTCTTCTCCACTAGAAATGAGAAAAACGATCAGAAATGAGCACAACGACTCTTAT 3924
Qy 1424 gaagaatggaactagatgattcttcgaagattacccttttgcgagtctcctagcagt 1483
Db 3923 GAAGAATGAACTAGAGATGGATCTTCGAAAAGTACCCTTTTGGGATGCTCTAGCAGT 3864
Qy 1484 tctgatttttaatccgattg----- 1503
Db 3863 TCTGATTTTAAATCCGATTTGGGTAAAGCTTCTATTCATCTGAAAAATAGCGTTAAATCCTT 3804
Qy 1504 -----gattgctcgtggaagtgcgatatcttcaaaagccgctgcagaagctccgattgc 1558
Db 3803 GTTTAGATTGCTCGCTGGAAGTGGGATATTCTCAAAAGCCCGCTGCAGAAAGTCCGATTGC 3744
Qy 1559 ctccccgttcgagcatggaagatgattgtagccccgagatggaactagcactcgacgct 1618
Db 3743 CTCCCCGTTCGAGCATGGAAGATGATTGATGACCCGGATGGAAC TAGCACTCGACGCT 3684
Qy 1619 tttctggaaggagatcatcaatatgactatgtctggtgttcaacatcttaagtat 1678
Db 3683 TTCTCGGAAGGAGTATCATCAATATGAGCTATGCTCTGGGTGTTCAACATCTTAATGAT 3624
```

Qy 1679 cataatgtggttctcaaaactgtgatccatggtgagccctgttcaagacttcacgtccgt 1738
|||||
Db 3623 CATATATGTGGTGTCAAACTGCTGATCCATGGTGACCCGTTCACAGACTTCATGTCGGT 3564

Qy 1739 ttcatggcagacttttctgacgactcagagaagcagagccaggttgaaactctgaaa 1798
|||||
Db 3563 TTCAATGGCAGACTTTTGTGACCACTCGAGAAAGCGGAGAGCCGAGTTGAACCTCGGAAA 3504

Qy 1799 tttaagaagatctcagagaaagtctcgagatgtctctgcaacgcttgagatcgatcc 1858
|||||
Db 3503 TTTGAAAAGATGCTCAGAGAAAGTTCTCGGAGTGCTTTGCAACGTTGGATCGATCGCTCC 3444

Qy 1859 atcaccgggggttattcggtgttttcggttggctgggaatgcgtctgacatcttttga 1918
|||||
Db 3443 ATCACCGGGGGTGTATTCGGTGTTCGGTTCGGTGGGAATGCGTTCACATCTTTTGAA 3384

Qy 1919 ttgdtgtgatacgggagatacatcgaagaagcgcaggtccaccagagcctgtctc 1978
|||||
Db 3383 TTGGTTGTGGATCGGGAGATACATCGCAAGAAAGCGCAGGTCCACCACGAAGCGCTGCTC 3324

Qy 1979 agtcgtttgttaggagtcacgcgagactgcagttctctctatcatgaaattcat 2030
|||||
Db 3323 AGTCGTTGTGAGGAGTCATCGCAGACTGCGAGTTCTCTATCATCAAAATTCATCAGGTTTA 3264

Qy 2031 ----- 2030

Db 3263 CTGTCAAAATCCAAATTTTGCATTCAAAATATATATTACTAGAGTAATCCTACAGTAC 3204

Qy 2031 ----- 2030

Db 3203 TGCCTACAGTACCCTACAGTACTACTACAGTACCCTGACCATATCCCCCACTAACTCAAA 3144

Qy 2031 ----- 2030

Db 3143 CTAATATCCCTTCAAAAGACAAAAGTCAATTTTCCAAAACACTACAGTAATCCTTCAGTA 3084

Qy 2031 ----- 2030

Db 3083 CTCCTACGGTACCATTACAGTACTACTACAGTACCCTGACCATATCCCCCACTAACTCC 3024

Qy 2031 ----- 2030

Db 3023 GAACTAATAATCCCTTCAAGAGACAAAACATGTATTTTATGTATAAGTAATGATTTT 2964

Qy 2031 -caactcatctaaggtatcactggaacttcgaagacactatgaaccatccgcct 2089
|||||
Db 2963 CCAGCTCCATCTAATGGGTATCACTGGAAACTTCGAAGACACCTATGAACCATCCGCCCT 2904

Qy 2090 aacgggctcttcatgtccctctgtgcagtaaaccttgcgaagctgcgcgagcatcaaa 2149
|||||
Db 2903 AACGGGCTCTTCATGTCCCTCTGTGCAGTAACCTTGCTGAAAGCTGCCGGAGCATCAA 2844

Qy 2150 cgagcacttcacgcgcgcctcatggttcagatcatatcttctgcattccatcccaatgccg 2209
|||||
Db 2843 CGACGGACTTCCACGCGCGCGCATGGCTCAGATCTACATTTCTGCATCCATCCAATGCCG 2784

Qy 2210 tttagctcttcgaacactactcgcacattctctcgggatacctttttacgaagagctcg 2269
|||||
Db 2783 TTTGGCTCTTCGAACCTACTCCGACCATTTCTCTCGGGATACCTTTTACGAAGAGCTCG 2724

Qy 2270 aagcacgtgcgtcagctccggagcactcggtgtccatttctgtatggatcttccatcc 2329
|||||
Db 2723 AAGCACGTGGCTCGAGCTCCGGAGCACTCGGTGTCCATTTGTTATGGATCTTCCATCC 2664

Qy 2330 agcacaagaagtctcatgtcagatcggaaggttgagcatgtgtgagctcgaagca 2389
|||||
Db 2663 AGCCACAAGAAGTTTCATGTTCAGATGCGAAAGGTTGGAGCATGTGTGAGCTCGAAGCA 2604

Qy 2390 gaacagttgagatttggggtcttttgggaagatgagcaat- 2430
|||||
Db 2603 GAACAGTTGAGATTGGGTCTTTTGTGGAAGATGAGCAATGTAAGTATTTTAAAAATAAA 2544

Qy 2431 -----latecccaacttgcctgaatc 2450
|||||
Db 2543 ATAAAAATAAAAAATAAATAAATTTTCATTTTCCAGTATCCCCACTTGCCTCGAATC 2484

Qy 2451 cgaacaacgctgaaagtgtacctactctccaaacttttacaggaaacttgcgtggtgac 2510
|||||
Db 2483 CGAACACGCTGAAAGTGTACTCTCCAAACTTGTACAGGAACTTGTCCGTGCTGAC 2424

Qy 2511 gagatcttttcaaaaaatgtgaaacgcatcctctaaatgacatgaccgtctcgnatatgaa 2570
|||||
Db 2423 GAGATCTTTTCAAAAAATGTGGAAGCATCTCTAAATGACAAATGACCGTCTCGATGATGAA 2364

Qy 2571 gtacacgtggtgtgatttcaagacttttggtagaatttcaacgcagctgcgctgccatt 2630
|||||
Db 2363 GTAGACGTGGTTGATGTTTCAAGACTTTTGGTGACAAATTTCAACGCAGTCCGCTGCCATT 2304

Qy 2631 ttgactaatgagaaggtgagtcagcga- ----- 2658
|||||
Db 2303 TTGACTAATGAAAGGATGAGTCAAGCGAGTTAGTGAACCGCAATTTTAAGATTTGAAAAATG 2244

Qy 2659 -----aattcgaacactgagatctctcgaacggaatgcttcttaccatg 2704
|||||
Db 2243 TTATAAAATTTCTAGAAATTCGGAACCTGGATCTCTCGAACCGAGATGCTTGTTCACATG 2184

Qy 2705 gtggacgcacgtcttgacatgtggaatctattggaggagtaacaagaatgcagctgcgcag 2764
|||||
Db 2183 GTGGACGCACGTTCTGACATCTGGAATCTATTGGAGGAGTAAACAGAGATGAGTGGCAGC 2124

Qy 2765 gcaacactattcactgacgaaactgcgcgcgaagatatttgacaga 2812
|||||
Db 2123 GCAACACTATTTCAGTGTAGCAACTGTCCGCCGAAGATTTTGACAGA 2076

RESULT 3
CEY47D3/c
LOCUS
DEFINITION
IN PROGRESS ***, in unordered pieces.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
Location/Qualifiers
1. .337565
/organism="Caenorhabditis elegans"
/db_xref="taxon:5239"
/chromosome="III"
/clone="Y47D3"
BASE COUNT 108087 a 59856 c 60930 g 107892 t 800 others
ORIGIN

```
Query Match          29.5%; Score 1008.4; DB 31; Length 337565;
Best Local Similarity 76.6%; Pred. No. 2.1e-280;
Matches 1508; Conservative 0; Mismatches 1; Indels 459; Gaps 4;

QY 1304 attaccgtaccggaaccattcaatacactgaataactctgcccgcatacccgctgaatc 1363
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304686 ATTACAGTACC CGGAACCAATTAATACACTGAATACTCTGCCGATACCCGTCGAATC 304627

QY 1364 atctctctccacttagaataagagaaacagatcacgaatgagacaaacactcctat 1423
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304626 ATCTCCTCTCCACCTAGAAATGAGAGAAACGATCACGAATGAGCACAACGACTCCTAT 304567

QY 1424 gaagaatgaactagatggtcttcgaagtaccctttttgcatctcttagcagt 1483
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304566 GAAGAATGAAC TAGAGATGGATCTTCGAAGATTACCCCTTTTGGCATCTCTTAGCAGT 304507

QY 1484 tctgatttttaattccgattg----- 1503
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304506 TCTGATTTTAA TCCGATTGGGTAAAGCTTCTATTTCATCTGAAAAATAGCGTTAAATCCTT 304447

QY 1504 -----gattgctcgtggaagtgcgatatctcctaaagcgcgtgcgaaggtcccgattgc 1558
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304446 GTTTAGATTGCTGCTGGAAGTGCATATTTCTAAAAGCCGCTGCAGAAAGCTCCGATTGC 304387

QY 1559 ctccccgttcgagcatggaagatgattgatgacccggatggaactagcactcggacgct 1618
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304386 CTCCCCGTTTCGAGCATGGAAGAGTATTGATGACCCCGGATGGNACTAGCACTCGGACGCT 304327

QY 1619 ttctcgggaaggagtatcatcaatatgagctatgctcgggtgttcaacatcttaatsat 1678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304326 TTTCTCGGAAGGAGTATCATCAATATGAGTATGCTGGTGTTCACATCTTAATGAT 304267

QY 1679 catatatgtggtgtcaaacactcgtatccatggtgacctgttcaagaacttcatgtccgt 1738
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304266 CATATATGTTGTTGCAAACTGCTGATCCATGCTGATCCCTGTTCAAAGACTTCATGTCCT 304207

QY 1739 tcatgacagacttttgacactcgagagaagcgagcgaagtgaactctggaaa 1798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304206 TTCATCGCAGACTTTTGTGACGACTCGAGAAAGCGCAGAGCGGATTTGAACCTCGGAAA 304147

QY 1799 ttgaaagatgctcagagaaattctcgcagtgcttgcgaagtgttgcgaagtgcgactcc 1858
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304146 TTTGAAAGATGCTACAGAAAGTTCTGCGAGTCTCTTGCAACGTTGGATCGATCGCTTCC 304087

QY 1859 atcacgggggttgattcgggttttcggttggtggaatgcggttcgcacatcttttgaa 1918
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304086 ATCACGGGGGTTGATTTCGGTGTCTTTTCGGTGGGAATGCGTTTCGACATCTTTTGAA 304027

QY 1919 ttggtttggaatcggaatatacatcgcaagaagcgaggtccacacgaagcgtctctc 1978
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304026 TTGGTTGTGGATCGGGAGATACATCCCAAGAGGCGCAGTCCACCACCAAGCCCTGCTC 303967

QY 1979 agtcgtttgtagagtcatagcgcagactgcagttctctatcatgaataatcat 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303966 AGTCGTTTGTAGGAGTCATGCGCAGACTGCAGTTCTCTATCATGAAATTCATCAGGTTTA 303907

QY 2031 ----- 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303906 CTGTCAAAATTCATAAATTTTGCATTTCAAAAATATATTTACTAGAGTAATCCTACAGTAC 303847

QY 2031 ----- 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303846 TGCTACAGTACCCCTACACTACTACTACAGTACCCTGACCATATCCCCTAACTCCAAA 303787

QY 2031 ----- 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303786 CTAATATCCCTTCAAAAGACAAAAAGTCAATTTTCCAAAACTACAGTAATCCTTCAGTA 303727

QY 2031 ----- 2030
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303726 CTCCTACGGTACCAATTACAGTACTACTACAGTACCGTGAACCATATCCCCCCTAACTCC 303667
```

RESULT 4
CEH10N23/c
LOCUS
DEFINITION
IN PROGRESS ***, in unordered pieces.
ACCESSION
VERSION Z94157.1 GI:3377969
KEYWORDS HTG: HTGS_PHASE1.

CEH10N23 36780 bp DNA HTG 22-SEP-1998
Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
Z94157
VERSION Z94157.1 GI:3377969
KEYWORDS HTG: HTGS_PHASE1.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 36780)

Sulston,J.

Direct Submission

Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu

On Aug 3, 1998 this sequence version replaced gi:1945146.

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

source

BASE COUNT

ORIGIN

Location/Qualifiers

1..36780

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="H10N23"

8697 a 5380 c 5339 g 8564 t 8800 others

Query Match

Best Local Similarity

Matches

661;

Conservative

0;

Mismatches

2;

Indels

45;

Gaps

1;

Qy

1304

attaccgtaccggaaacaaattcaatacactgaatacttccccgatacccccgtcgaaac

1363

Db

708

ATTACAGTACCGGAAACCAATTCAATACACTGAATACACTCTCCCGGATCACCCGTCGAATC

649

Qy

1364

atctccttctccactagaaatgagaaaaacgacatcacgaatgagcaaacgactcttat

1423

Db

648

ATCTCTCTCTCCACCTAGAAATGAGAAAAACGATACAGAAATGAGCAACAGCATCTCTAT

589

Qy

1424

gaagaatggaactagatggtattctgaaagtttaaccttttttgcgattgctccctagcagt

1483

Db

588

GAGAATGGAAGTACAGATGGATCTTCGAAAGTTACCCCTTTTTCGGATGCTCTTAGCAGT

529

Qy

1484

tctgatttttaatccgattg-----

1503

Db

528

TCTGATTTTAATCCGATTGGGTAAAGCTTCTATTTCATCTGAAAAATACCGTTAAATCCTT

469

Qy

1504

-----gattgctcgtggaatgcgattctctcaaaagccgctcagaagctccgattgc

1558

Db

468

GTTTAGATTGCTCTCGTGAAGTGCATATTCTCAAAAGCCCTCTCAGAAGCTCCGATTGC

409

Qy

1559

ctccccgttcagcatggaagagtgattgatgaccggatggaactagcactcgacgct

1618

Db

408

CTCCCCGTTTCAGGACATGGAAGAGTGATTGATGACCCCGATGGAACTAGCACTCGGACGCT

349

Qy

1619

ttctgggaagggagtgatcatcaatatgagctatgtctgggtgttcaacatctttaatgat

1678

Db

348

TTTCTGGGAAGGGAGTATCATCAATATAGCTATGTCTGGGTGTCAACATCTTAATGAT

289

Qy

1679

cataatgtggttgcbaaactgctgatccatggtgagccctgtttcaagacttcattgctcgt

1738

Db

288

CATATATGTGTTTGTCAAACCTGCTGATCCATGTTGATGTCGACCTGTTCGAAGACTTTCATGTCGT

229

Qy

1739

ttcatggcagacttttctgacgactcgagagaagcgagagccgagtgtaactctgaaa

1798

Db

228

TTCATGCGACACTTTTGTGACGACTCGAGAGAGGGCGAGAGCCGAGTTGAATTTCTGAAA

169

Qy

1799

tttgaagaatgctcagagaaagtcttcgcgagtgcttgcacagttggatcgatcgcttcc

1858

Db

168

TTTTAAAGATCTCTCAGACAAAGTTCTCGAGTGTCTTCAACGTTGGATCGATCGCTTCC

109

Qy

1859

atcaccgggggttgatttcctcggtgttttcggttggctgggaatgcgttcacatcttttga

1918

Db

108

ATCACCGGGGTTGATTCCGGTGTATTTCGGTTGGCTTGGGAATCGTTCGCACATCTTTTGA

49

Qy

1919

ttggttggtatcgggagatatacatcgcaagaagcgccaggtccaccac

1966

Db

48

TTGGTTGTGGATCGGAGATACATCGCAAGAAGAGCGCAGGTCCACCAC

1

RESULT

5

CEY48A6

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CEY48A6

296699 bp

DNA

Caenorhabditis elegans chromosome III clone Y48A6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.

Z92854

GI:3218061

HTG; HTGS, PHASE1.

Caenorhabditis elegans.

Caenorhabditis elegans

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1

(bases 1 to 296699)

Direct Submission

Submitted (30-JUL-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu

On Jun 13, 1998 this sequence version replaced gi:3204147.

Order of segments is not known; 800 n's separate segments.

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

source

BASE COUNT

ORIGIN

Location/Qualifiers

1..296699

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="Y48A6"

93035 a 52118 c 54334 g 90807 t 6405 others

Query Match

Best Local Similarity

Matches

614;

Conservative

1;

Mismatches

6;

Indels

56;

Gaps

2;

Qy

2799

aagattttgacagacaatctcgttttgccggttgcgcacgcggttgctgcgaagatt

2858

Db

19474

AACATTTTTTCCAGCAATCTCGGTTTGCGGTTTGCCACCGCGTTGTGTGCTCGCAAGATT

19533

Qy

2859

tgcataatgacgagatctccccaagtcagtcacatcagtcggtgcttcacacaaaaga

2918

Db

19534

TGCATAGATGACCGAGATTCCTCCCGAAAGTCAGTCAATACGTTGCATTTCACACAAAAG

19593

Qy

2919

tcgctgaatccctccgactattctccacatcatcgcgagcatcaggtgtggtctcga

2978

Db

19594

TCGCTCCCAATCCCTCCCACTATTCTCCACATCATTCGGGAGCATCAGGTGTGCTGGA

19653

Qy

2979

attcagaagtgacacgccgaatggcctacgaatgattgaactcgtcgtcgacgcg

3038

Db

19654

ATTCAGGAAGGTACACGCCGAATGCGCTACGAATGGATTATGAACCTCGCTCGACGCG

19713

```
QY 3039 tggcggtccaattatctgcacatgaacccctactggacacaaagcttcaaggagacaatcc 3098
|||||
Db 19714 TGGCGTTCGAATCTATTCGCATCGAACCCTACTGGACACAAAGCTTCAAGGACAATCC 19773
|||||
QY 3099 acggttagtcacgtttatcaagaaggcgtataatcaatcattatgcgattattaatgggacaag 3158
|||||
Db 19774 ACGTTTGTAGTACGCTTATCAAGAGGCGTATAATCATATATGCGATTATTAAATGGGACAAGG 19833
|||||
QY 3159 ggagattgttgaga----- 3173
|||||
Db 19834 GGAGATTGTTGAGAGCTGAGTTATTACATTACTCTCCAGTAACCTAAACTAAACTCTTT 19893
|||||
QY 3174 -----ctattgtctagagctcagtgccgaatgctcaacggagcccaaccaca 3223
|||||
Db 19894 ATCGTTTTAGCTATTGTCTACGAGCTCAGTGCAGGAATGCTCAACGGAGCAACCCACA 19953
|||||
QY 3224 agccagctggtcagggcgcgcgctcgatctacaaaaatggagcgcggtccggagag 3283
|||||
Db 19954 AGCCAGTGGTCAGGCGTCCGAGCGCTTCGATCTACAAAAATGGAGCGGTCGGAGGAAG 20013
|||||
QY 3284 agtgaacatgc-gacgctcggtccaacggacgcgcatcttcatcttcatacactgggttaaac 3342
|||||
Db 20014 ASTGAGCATGCGGACGCTCGGCTCAACCGGACGCAFTTTCATCTTCATACACTGGTTAAAC 20073
|||||
QY 3343 tacatactctatggatcttgaattgaacaaaaatgattttattcagaataaataatgataa 3402
|||||
Db 20074 TACATACTTCTATGGATCTTGAATTTGACAAAAAATGATTTATTCAAGTAATGATTA 20133
|||||
QY 3403 atacgattatatataa 3419
|||||
Db 20134 ATACGATTATATATAA 20150
|||||
```

```
RESULT 6
CEH10N23 36780 bp DNA HTG 22-SEP-1998
LOCUS Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
DEFINITION IN PROGRESS ***, in unordered pieces.
ACCESSION Z94157
VERSION Z94157.1 GI:3377969
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
TITLE 1 (bases 1 to 36780)
JOURNAL Sulston,J.
COMMENT Direct Submission
Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
On Aug 3, 1998 this sequence version replaced gi:1945146.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
FEATURES
source
Location/Qualifiers
1..36780
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="H10N23"
BASE COUNT 8697 a 5380 c 5339 g 8564 t 8800 others
ORIGIN
```

```
Query Match 12.4%; Score 424.2; DB 31; Length 36780;
Best Local Similarity 86.8%; Pred. No. 3.2e-111;
Matches 501; Conservative 0; Mismatches 23; Indels 53; Gaps 1;

QY 544 agaatttcactcatgaatgctatcttcaccacacccatcctcaatgacttctcaacct 603
|||||
Db 5925 AGAATTTCACTCATGCAATGCTATCTCCACCACACCATACCTCAATGACTCTCTCAACCAT 5984
|||||
QY 604 atacagaagccatgggacatatcaacgggtacatgtctccatacagacca----- 652
|||||
Db 5985 ATACAGAAGCAATGGAACATATACACGGGTACATGTCTCCATACGACCAAGGTTTTCTTT 6044
|||||
QY 653 -----agctcaaggcccatcag 670
|||||
Db 6045 TCCCTTTTACTGATTAACCTTATTATGATCCTTAATAATTTCCAGCTCAAGGCCCATCAGG 6104
|||||
QY 671 accatcatattactcacacacacacacacacacacacacacacacacacacacacacac 730
|||||
Db 6105 ACCATCATATTACTCACACACCATCAATCTCCACCACCTCATCACCAACCATCACCAACC 6164
|||||
QY 731 gatgcaaaaaatccatgagaaacctgaaacagtggcatctccatcctcatgattgagatgctcc 790
|||||
Db 6165 GATGCAAAAAATCCATGAGAACCCCTGAACAAGTGGCATCTCCATCGATTGAAGATGCTCC 6224
|||||
QY 791 agagacgaacacacacacacacacacacacacacacacacacacacacacacacacacac 850
|||||
Db 6225 AGAGACGAACCAACTCATTTGTTGAACCAACANAGTCCAAAAAGCCCGCAGCAATATGAA 6284
|||||
QY 851 agagagctctctcggttactagttaacatgtctctccagtgaggttgaacggttaagaa 910
|||||
Db 6285 AGAGGAGCTTCTCGGTTACTAGTTAAACATGTCTCCAGTGAAGTTGAACGGTTAAAGAA 6344
|||||
QY 911 taaaaaacaggagagcatgttcagcgacgaatgggcccacagaggagtaagagagaagcgcc 970
|||||
Db 6345 TAAAAAATCAGAGCATGTTCAGCGACGAATGGGCATCGAGGAGTAAGGAGAGGCGGC 6404
|||||
QY 971 gaagattgtatcaggagacagcggaaggatgaagatgaggaatgaggaatgaggaatgga 1030
|||||
Db 6405 GAAGATTGTGATTCAGGACAGCCGCAAGGGCATGAAGATGAGGATGATGAGGATAGTCA 6464
|||||
QY 1031 ttccgggggagactatgtctcagggaactactattatt 1067
|||||
Db 6465 TTACGCACACACTATCACTGCGAACTGACTCATATT 6501
|||||

RESULT 7
I66494/c 7218 bp DNA PAT 28-DEC-1997
LOCUS Sequence 14 from patent US 5670367.
DEFINITION I66494
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source
1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 2.1%; Score 71.8; DB 5; Length 7218;
Best Local Similarity 11.7%; Pred. No. 3.4e-09;
Matches 58; Conservative 231; Mismatches 208; Indels 0; Gaps 0;

QY 771 ccacgattgaagatgctccagagacgaacacacactcatttggttgaaccacaaagtcca 830
```

```

Db 1475 CTATCTATGCAAGTAGTAAAGAGATAGAAGATTTCGTACRRRRRRRRRRRRRR 1416
QY 831 aaagcccgagaaatagaagagagcttcttcggttactagttaacatgtccagat 890
Db 1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356
QY 891 gaagtgaacggttaagaataaaaaatcaggagcatgttcagcgagaaatgggcccgcg 950
Db 1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296
QY 951 agagtaagagagagcgcggaagattgtattcagagagacgagcggaagggatgaagat 1010
Db 1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236
QY 1011 gagatgatcagatagtattccggggagaaactatgtctcaggaaactattattggt 1070
Db 1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176
QY 1071 cgaagaccaaaacccgagcgtcgtacggcacacaaatctcatcgaagaagatagatgc 1130
Db 1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116
QY 1131 tcaataatgatgaattcaacagctgaagtagtactttgtgtgggagatgaagcttaagctt 1190
Db 1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1056
QY 1191 tcaaaatcgcacactacgacgagcgtattgaacatcatcgaggaggttgaacagagaat 1250
Db 1055 CCTGACCTCGACCAAGCTCGGAATTAATCTGTGAGCGTATGCGCAACGAGGAAAA 996
QY 1251 caggtgttgaagcatca 1267
Db 995 TAGTTATAGTAGCGCA 979

RESULT 8
CGU09103 3816 bp mRNA ROD 26-JUL-1994
LOCUS Cricetulus griseus sterol regulatory element binding protein-1
DEFINITION (SREBP-1) mRNA, complete cds.
ACCESSION U09103
VERSION U09103.1 GI:516002
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa, Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
1 (bases 1 to 3816)
Sato,R., Yang,J., Wang,X., Evans,M.J., Ho,Y.K., Goldstein,J.L. and
Brown,M.S.
Assignment of the membrane attachment, DNA binding, and
transcriptional activation domains of sterol regulatory binding
protein-1 (SREBP-1)
J. Biol. Chem. 269, 17267-17273 (1994)
94274723
REFERENCE 2 (bases 1 to 3816)
AUTHORS Yang,J.
TITLE Direct Submission
JOURNAL Submitted (20-APR-1994) Jianxin Yang, Molecular Genetics,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd., Dallas, TX 75237-9046, USA
FEATURES
source 1..3816
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/cell_line="CHO"
/cell_type="CHO K1"
31..3432
/feature="SREBP-1"
/codon_start=1
/product="sterol regulatory element binding protein-1"
/protein_id="AAA20085.1"
CDS

```

```

/db_xref="GI:516003"
/translation="MDELPTGAAVEQALDELGDAALLTDIDQMLQINNODSEFP
GLFDSVYAGGAGDTPTSPANCSPESLSASLGSSLEAFLOECPKATPSLSVPVA
STALKWYSPVPPSPGGIKEEVPPLTILOFPAAOPSGHTLLPSPFPPLQISPAVP
LGYSSLPSPSGTLPNGTQPPSSLSASAPGVSP1SLHTQVSSASQOPLPASTAPR
TLVTISQIRPVVLPQHFHKADSLTLTYKTDTGTATKTAGTSLAGTAVOAGPLQ
DLVSGGTILATVPLVVDTKLPIHRLAAGSKALGSAQSGEKRTHAHNAIKRVRSSIN
DKIVELKDLVVGTEAKLNKSAVLKATDYIRFLQHSNOKLKOENLARNAHAKSKLK
DLVYSCSAGGTDVAMEGVKPEVVDILTTPSDAGSPSSPLSLSGSGSSGSDSE
PDSVPFEDSOVKAORLHSHGMRLDRSLALCALVPLCTICNPLSLFWGVIPIGSSASG
AHHSSRMLEAESRDGNSNTOWLLPVLWLANGLLVLACLALFLVYGEVPTHTSP
AFVHRRKOADLDLARGDFAAOOLWLALQALGRPLPTSLNDLACSLIWNLTIRHL
QRLVWGRWLAGRAGGLRDCGLRMDARASDAALVYHKLHQLHAMGKYTCGHLIASN
LALSALINAPCADGASMTAEIYVAALRVKTSYLPRAHLFTLFTLSSARQACLAQ
SGSVPLAMOWLCHPVGHRTFVDWAVIHGAPOESLYSVAGNPVDPVLAQVTKLCEHLL
ERALNCIAOPSGTADGDRFSDALCYLLOLNRCSDAVGTACFSFVSSSMASTTGTD
PVAKWASLTAVIHWLRDDEAAERLYPLVERPHVLOETRPLPKAALYSKAAARA
LDHRRVSEGPASLTACEKASGLRDSLAAPPTGSSIDKAMQLLCLLDLLIVARTSMWO
ROQSPASQAQVHASNGSOASALELRGFQODLSLRRLAONFRPAMRRVFLHEATRL
MAGASPARTHOLLDRSLRRRAGSGKGTVAELEPRPTWREHTEALLLASCYLPPAFL
SAPGOOMSLAEAAARTVEKLGDRHLLDDCCQMLRLGLGGTIVTSS"
polyA_signal 3793..3799
polyA_site 3816
BASE COUNT 770 a 1188 c 1061 g 797 t
ORIGIN
Query Match 1.7%; Score 57.4; DB 12; Length 3816;
Best Local Similarity 53.3%; Pred. No. 4.5e-05;
Matches 121; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 1086 gacgctcgtacgcacacaaatctcatcgaagaagatatagatcctaataatgacga 1145
Db 979 GAGAGCGCAGACCCACAAATGCCATTGAGAAAGCCCTACCGCTTCTCTATCAATGACAAG 1038
QY 1146 attcaacagctgaaagtacttttgtgtgggagatgaagcttaagcttcaaaatcggaaca 1205
Db 1039 ATTGTGGAGGCTCAAGACCTGGTGGGCACTGAGGCAAGACTCAATAATCTGCCGTC 1098
QY 1206 ctacacgagctattgaacatatcagagaggttgaacacgagaaatcagagtttaagacat 1265
Db 1099 TTGCCAAGGCCATCGACTATATCCGCTTCTTACAGCAGCAGCAGCAGCAAGCTCAAGCAG 1158
QY 1266 catgttgacaaatgaadaacacactcgaataatcgattaccgta 1312
Db 1159 GAGAACCTGGCCCTCGGAATGCGGCTCACAAAAGCAAAATCCCTGAA 1205

RESULT 9
AB017337 1209 bp mRNA ROD 14-APR-1999
LOCUS Mus musculus mRNA for sterol regulatory element-binding protein-1
DEFINITION (SREBP-1), partial cds.
ACCESSION AB017337
VERSION AB017337.1 GI:4240011
KEYWORDS sterol regulatory element-binding protein-1 (SREBP-1).
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE Inoue,J. and Sato,R.
AUTHORS A novel splicing isoform of mouse sterol regulatory element-binding
TITLE protein-1 (SREBP-1)
JOURNAL Biosci. Biotechnol. Biochem. 63 (1), 243-245 (1999)
MEDLINE 99161303
REFERENCE 2 (bases 1 to 1209)
AUTHORS Inoue,J. and Sato,R.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1998) to the DDBJ/EMBL/GenBank databases. Jun
Inoue, Graduate School of Pharmaceutical Sciences, Osaka
University, Department of Biochemistry and Molecular Biology;

```


JOURNAL	Submitted (15-JUL-1994) Jianxin Yang, Molecular Genetics, Univ., Dallas, TX 75235-9046, USA										
FEATURES	Location/Qualifiers										
source	1..1914 /organism="Cricetulus griseus" /db_xref="taxon:10029" /clone="PMThamSRBP-2" /cell_line="SRD-1" /cell_type="hybrid Chinese hamster ovarian cells" /note="hybrid CHO cells include a tranlocation of the SRBP-2 gene"										
gene	155..1570 /gene="mutant SRBP-2"										
CDS	155..1570 /gene="mutant SRBP-2" /codon_start=1 /product="mutant sterol regulatory element binding protein-2" /protein_id="AA74140.1" /db_xref="GI:551504" /translation="MPDESSELGLTMDTLTELGDLELTIDEMLQFVSNQVGEPFD LFSEOLCSFFGGSGSSGSSNSNGGAATPCGVORFSQVPLPTSP STASPALALOVKYVSTPPRATPVLPQPPOPOPQSALOQOVTWITPFTSTAOT RIQQPLITONATSFQVLQAQSVLSVSQOPVTIQOOQTVAQRVLTTGTANGTL QLAPATAVTAAPQVOGPVLVPPQI IKTDLSLLTKLDGSPVMAAQNPAALT TFLTGALQVPTLVGSNGVILTPMVMGQEKVIPKQVPGVKOLEPPKGERRTTHN ITEKRVRSINDKIIEUKDLVMTDKAMHKSGVLRKAIDYIKYLOOVNHKLROENVVL KLANOKNKLLKGIDLGLSLVDSDVLDLKIEDFNQNVLLMSPPPASPSSQGAFSPYSIDUSE PGSLDDAKMKKVIOVKSPR"										
BASE COUNT	495 a 540 c 501 g 378 t										
ORIGIN											
Query Match	1.6%; Score 55.6; DB 12; Length 1914;										
Best Local Similarity	53.2%; Pred. No. 0.00013;										
Matches 118:	Conservative	0;	Mismatches	104;	Indels	0;	Gaps	0;			
QY	1079	aaaaaccgagcgtcgtagccacacaatctcatcgaaagaagtatagatgcgtcaataaa								1138	
Db	1129	AGAAGGAGGAGGAGGCACACACACACATCATGTGAAAACCGATATCGATCCCTCATCAA								1188	
QY	1139	tgatgaattcaaacagctgaagtacttttgttggggatgaagctaaggctttccaaaatc								1198	
Db	1189	TGCAAAATCATAGACTGAAGGACATTGGTCATGGGAACAACATGCCAAGATGCCAACATC								1248	
QY	1199	ggcaacctactcacggcgtattgaacatatcgaggaggtttaacacagagaatcagggttt								1258	
Db	1249	TGGCGTTCTGAGGAAGCCCAATTGATTACATCAATAATTTGCAGCAGGTCATCACAAAGT								1308	
QY	1259	gaagcatcatgtigaacaaatgagaagacacactgcadaataa								1300	
Db	1309	TCGTGAGGAAACATGGTACTGAGGCTGGCAAAATCAGAAAAA								1350	
RESULT 12											
CGU22818											
LOCUS	CGU22818 4425 bp mRNA ROD 26-JAN-1996										
DEFINITION	Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SRBP-2) mRNA, complete cds.										
ACCESSION	U22818										
VERSION	U22818.1										
KEYWORDS	GI:841317										
SOURCE	Chinese hamster.										
ORGANISM	Cricetulus griseus										
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.										
REFERENCE	1 (bases 1 to 4425)										
AUTHORS	Yang, J., Brown, M.S., Ho, Y.K. and Goldstein, J.L.										
TITLE	Three different rearrangements in a single intron truncate sterol regulatory element binding protein-2 and produce sterol-resistant phenotype in three cell lines. Role of introns in protein evolution J. Biol. Chem. 270 (20), 12152-12161 (1995)										
JOURNAL											

59263566
2 (bases 1 to 4425)
Yang, J.
Direct Submission
Submitted (14-MAR-1995) Jianxin Yang, Molecular Genetics,
University of Texas Southwestern Medical Center, 5323 Harry Hine
Blvd, Dallas, TX 75235, USA
Location/Qualifiers
1..4425
 /organism="Cricketulus griseus"
 /db_xref="taxon:10029"
 /cell_line="SRD-2"
 /size=138
 /gene="SREBP-2"
 /size=138
 /gene="SREBP-2"
 /codon_start=1
product="mutant sterol regulatory element binding
protein-2"
protein_id="AAA85718.1"
 /db_xref="GI:841318"
translation="MPDESSELGLETMDTLFELGDELITLGDIDEMLOFYSNVOYGEPEP
LFSEQLCSFPFGSGSSSSSSSSSSSSSCNSRNGCAATDPGVORFSOVPLPTSP
RIQSPALLALVKVSTPPRPATPVLPQPPOPOPVSOPQPTIQQOQOTVMITPTTETAPOT
RIIQPLLITYONASTSVOLQAQSILVTSSOQVPPTIQQOQOTVQAOEVLTOTANGTL
QTLPATQVTPAAPOVQVPLVQPOLIKTDSLVLTLLTKDGSPVMAAVONPALTAIT
TPLQTCALQVPTLVGSNGTILTMPVMGOEKVPKIVPGGVKOLEPPKEGERRTTHN
ILEKRVSSINDKIIIEKLDMCTDAKMHSQVLRAIDYIKYLOOVNHKLROENVWL
KLANQRNLKGIDGLSDVSLKIEDFNQNVLLMSPASUSGSAQGFSPYSIDSE
PUSPLLDIAKKQAPAKEICVCQKTIVYPMERILLANOQVFHISCFRCSCNNKLSIGT
IPNSHQEIYCVKPHFNOLFKNKYNGDEGFGRHPKHDLWANKSENEFTSERSPQSDPVG
IFHNSQVEDAPISKVGVLAANMEAKASPQREKDPAETKKLRIAMPPPAELGGSGGT
MEEGFKSGPKWPPEDETSKTEALCDVDLIDLKLRSSSLKEBSRPTVAASFRTSL
KSFKTLSPLRKWSMSQNEESCGGIVTAGQTNENAKPKCKNKVRKRSWORDEDSARE
EVARRSKDIHGVELESENFIENGANIYEKSDGLVOOSPLDPKPSPSWSGFVDVTTPTTEE
FTTONKSHDVGRFEYGVREILSVEPQIKRNRYIDEDEEE"

BASE COUNT 1236 a 1086 c 1101 g 1002 t

Query Match 1.6%; Score 55.6; DB 12; Length 4425;
Best Local Similarity 53.2%; Pred. No. 0.00015;
Matches 118; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 1079 aaaaaccgagcgtcgtagccacacaatctcatcgaaagaagatagatgcataaa 1138
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 AGAGAGACAGAGAGGACACACACACATCATTTGAACGCCGATATCCATCCTATCAA 1171

QY 1139 tgatgaattcaacagctgaaagtacttttgtgtgggatgaagctaagctttcaaatc 1198
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 TGACAAAATCATAGTTTGAGGACTTTGGTCATGGGAACACATGCCAAGATGCACAATC 1231

QY 1199 ggcaaacactcagcgggctattgaacatatcggaggaggttgaacacgagatacagtggtt 1258
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1232 TGGCGTTCTCAGGAAGGCCATTGATACATCAATAATTTTCGACGAGCTCAATCACAGCT 1291

QY 1259 gaagcatcatgttgaacaaatgacaaagacacctgcagataaa 1300
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1292 TCCTCAGGAGAACAATGCTACTGAAGCTGGCAATCAGAAAAA 1333

RESULT 13
GU12330
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Cricketulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.

Db 1193 ATCAATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAAGCTGAATAAATCTGCTGTC 1252
Qy 1206 ctacgacgggctatttgaacatatacgagagggttgacacagagaatcagggtgttgaagcat 1265
Db 1253 TTGCGCAAGGCCATCGACTACATTTCGCTTCTGCAACACAGCAACCAAGAACTCAAGCAG 1312
Qy 1266 catgttgaacaaatgagaaaagacactgcagaaataatcgattaccgta 1312
Db 1313 GAGAACCTTAAGTCTGCGCACTGCTGTCCACAAAAGCAAAATCTCTGAA 1359

Search completed: September 3, 2000, 23:42:35
Job time: 18375 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:57:10 ; Search time 183.89 Seconds
(without alignments)
2557.526 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 gggttaattaccacaagtttg.....taaatcagattatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	71.8	2.1	7218	1	US-08-232-463-14
2	51	1.5	4154	1	Sequence 14, Appl
3	51	1.5	4154	3	Sequence 37, Appl
4	47.6	1.4	5197	1	Sequence 37, Appl
5	47.6	1.4	5197	3	Sequence 53, Appl
6	37.4	1.1	1981	5	Sequence 53, Appl
7	37.4	1.1	2295	1	Sequence 1, Appl
8	37.4	1.1	2295	5	Sequence 3, Appl
9	37.4	1.1	2295	6	Sequence 3, Appl
10	37.4	1.1	4080	1	Sequence 3, Appl
11	37.4	1.1	4080	5	Sequence 1, Appl
12	37.4	1.1	4080	6	Sequence 1, Appl
13	36.4	1.1	1027	5	Sequence 1, Appl
14	36	1.1	10747	4	Sequence 19, Appl
15	36	1.1	10747	5	Sequence 1, Appl
16	36	1.1	10747	6	Sequence 1, Appl
17	36	1.1	24979	4	Sequence 1, Appl
18	36	1.1	24979	5	Sequence 3, Appl
19	36	1.1	24979	6	Sequence 3, Appl
20	35.8	1.0	7218	1	Sequence 14, Appl
21	34.4	1.0	2344	5	Sequence 2, Appl
22	34.4	1.0	53526	5	Sequence 2, Appl
23	34.4	1.0	53577	5	Sequence 1, Appl
24	33.4	1.0	3489	4	Sequence 1, Appl
25	33.4	1.0	32207	3	Sequence 20, Appl
26	32.8	1.0	2277	2	Sequence 2, Appl

27	32.8	1.0	2277	2	US-08-676-974-2	Sequence 2, Appl
28	32.8	1.0	2277	4	US-09-098-487-2	Sequence 2, Appl
29	32.8	1.0	2721	7	5215881-2	Patent No. 5215881
30	32.8	1.0	8438	1	US-07-945-283-1	Sequence 1, Appl
31	32.6	1.0	570	1	US-07-885-970A-10	Sequence 10, Appl
32	32.6	1.0	570	1	US-08-298-687A-10	Sequence 10, Appl
33	32.6	1.0	570	1	US-08-298-829-10	Sequence 10, Appl
c 34	32.6	1.0	609	1	US-08-530-797-9	Sequence 9, Appl
c 35	32.6	1.0	609	4	US-08-787-335-9	Sequence 9, Appl
c 36	32.6	1.0	3680	2	US-08-759-848-1	Sequence 1, Appl
c 37	32.6	1.0	3680	6	PCT-US95-09383-1	Sequence 1, Appl
38	32.4	0.9	512	2	US-08-460-739-2	Sequence 2, Appl
c 39	32.4	0.9	2188	1	US-07-865-662F-10	Sequence 10, Appl
40	32	0.9	2085	1	US-08-371-930-9	Sequence 9, Appl
41	32	0.9	2085	6	PCT-US94-01712-9	Sequence 9, Appl
42	32	0.9	2088	3	US-08-104-165-4	Sequence 4, Appl
43	32	0.9	2088	7	5218100-4	Patent No. 5218100
44	32	0.9	2256	4	US-08-422-333-1	Sequence 1, Appl
45	32	0.9	2256	7	5187153-1	Patent No. 5187153

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```
Query Match          2.1%  Score 71.8;  DB 1;  Length 7218;
Best Local Similarity 11.7%  Pred. No. 1.2e-11;
Matches 58;  Conservative 231;  Mismatches 208;  Indels 0;  Gaps 0;

Qy 771 ccattgattgaagtcctccagagacgaaacaaactatttgggtgaaccacaaagtcca 830
    ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1475 CTATCTATCAAGTAGTTAAAGAGATAGAAGAAATTTGGTACACACACACACACACAC 1416

Qy 831 aaaaacccgcagaaatataaagagagagcttctcgttactagttaacatgtctccgagt 890
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy 891 aaagttaacggttaaaataaaataaaatcagagacatgttcagcagcaaatgggccatcg 950
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy 951 agagtaaggaagcgcgcgaagattgtgattcaggagacgcgcgaaggggagatgaagat 1010
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy 1011 gaggtatgagatagtagtattccgggagacatgtctcagggaactactattattgtt 1070
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy 1071 cgaagacaaaacccgagcgtcgtacgcacacacacatctcatcgaagaaagtatagatgc 1130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy 1131 tcaataaatgatcgaattcaacagcgtgaagtaacttttgggtgggagatgaagctt 1190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1056

Qy 1191 tcaaatcggcaacactacgacggcgtattgaacatacagagaggttgaaacacagagaat 1250
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1055 COTCAGCTGCGCAAGCTCGGAATTAATCTGTGACGCTATGCGCAACGAGGAAAAA 996

Qy 1251 caggtgtgaagcatca 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 TAGTTATAGTACGCCGA 979

RESULT 2
US-08-131-365B-37
; Sequence 37, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131.365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.

Query Match          1.5%  Score 51;  DB 1;  Length 4154;
Best Local Similarity 51.5%  Pred. No. 2.4e-05;
Matches 117;  Conservative 0;  Mismatches 110;  Indels 0;  Gaps 0;

Qy 1086 gagcgtctagcggcacacaatctcatcgaagaaagatagatgctcaataaatgatcga 1145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1133 GAGAAAGCGCACAGCCACCAACGCCATTGAGAAGCGCTACCGGCTCCTCCATCAATGACAAA 1192

Qy 1146 attcaacagcgtgaagtaacttttgggtgggagatgaagcttaagctttcaaatcggaaca 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 ATCATTTGAGCTCAAGGATCTGGTGTGGGCACTGAGGCAAAAGCTGAATAAAATCTGCTGTC 1252

Qy 1206 ctacgacggcgtattgaacatacgcaggaggttgacacagagaatcagggtgttgagcat 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 TTGCCCAAGGCCATCGACTACATTCGCTTTCTGCAACACACAGCAACCAAGAACTCAAGCAG 1312

Qy 1266 catgttgacaataatgagaaagacactgcagaataatcattaccctga 1312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1313 GAGAACCTTAAGTCTGCCGCACTGCTGCCACAAAAGCAAAATCTCTGAA 1359

RESULT 3
US-08-668-123-37
; Sequence 37, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668.123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131.365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
```

```

; REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-131-365B-53

Query Match 1.4%; Score 47.6; DB 1: Length 5197;
Best Local Similarity 54.6%; Pred. No. 0.00032;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 1079 aaaaacgagcgtgtagcgacacaaatctcatcgaaaagagtataatgatgtcctaataa 1138
Db 1098 AGAGGAGAAAGGCGACAAACCCATAATCATTCAGAAACGATATCGCTCTCCATCAAA 1157

Qy 1139 tgatcgaattcaacagctgaaatgacttltgtgtgggatgaagctaaagcttcaaaatc 1198
Db 1158 TCACAAAATCATCGAATTTGAAGAGCCTGGTCATGGGGACAGACGCCAAGATGCACAAGTC 1217

Qy 1199 ggaacactacgagggcgtattgaacatcatcgaggaggttgaacacgagaatca 1252
Db 1218 TGGCGTTCTGAGGAAGGCCATTGATTACATAATCTTGCAGCAGGTCAAATCA 1271

RESULT 5
US-08-668-123-53
; Sequence 53 Application US/08668123
; Patent No. 5891831
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:

```

Qy	889	gtgaagtgtgaacggtttaagaataataaaaatcaggagcatgttcagcgacacgaatggcccat	948
Db	1589	GTGTCTGATTACTAGGAGAAGATGAAGCGCGAGATAAGTTACACCGCAACACAGAGTCAG	1648
Qy	949	cgaaggagttaagagaaagcgcgcaagattgtgattctcaggagacacgacggaaggggatgaag	1008
Db	1649	CGCCAGGAANAACATCAGCGAAAGCAAGCAAAAGTGAAGATGAAGACGATGAGGAGCAGATG	1708
Qy	1009	atgaggatgatgaggatagtgattccgggga	1039

[illegible]

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. P.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/046001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)542-5070
TELEFAX: (617)542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-375-300-1

Query Match 1.1%; Score 37.4; DB 1; Length 4080;
Best Local Similarity 53.0%; Pred. No. 0.39;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 889 gtgaagtgaacggttaagaataaaatcaggagcatgttcagcgacgaatggcccat 948
DB 3038 GTGTCGATTACTAGGAGAAGATGAAGCGCGAGATAGTACACCGAACACAGAGTCAG 3097
QY 949 cgaggagtaaggaagcgcggaagattgtgtattcaggagacagcgggaagggatgaag 1008
DB 3098 GCCAGGAAAACATCAGCAAGCAAGCAAGTGAAGTGAAGATGAAGACGATGAGGACGATG 3157
QY 1009 atgaagatgatgatgatgtatttcgggga 1039
DB 3158 ACGAGATGATGACGATGACGATGACGATGA 3188

RESULT 11
US-09-177-431-1
Sequence 1, Application US/09177431
Patent No. 6071700
GENERAL INFORMATION:
APPLICANT: He, Feng
APPLICANT: Jacobson, Allan S.
TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
TITLE OF INVENTION: ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
FILING DATE: US/09/177.431
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/955,472
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter

REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/050001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-9806
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-177-431-1

Query Match 1.1%; Score 37.4; DB 5; Length 4080;
Best Local Similarity 53.0%; Pred. No. 0.39;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 889 gtgaagtgaacggttaagaataaaatcaggagcatgttcagcgacgaatggcccat 948
DB 3038 GTGTCGATTACTAGGAGAAGATGAAGCGCGAGATAGTACACCGAACACAGAGTCAG 3097
QY 949 cgaggagtaaggaagcgcggaagattgtgtattcaggagacagcgggaagggatgaag 1008
DB 3098 GCCAGGAAAACATCAGCAAGCAAGCAAGTGAAGTGAAGATGAAGACGATGAGGACGATG 3157
QY 1009 atgaagatgatgatgatgtatttcgggga 1039
DB 3158 ACGAGATGATGACGATGACGATGACGATGA 3188

RESULT 12
PCT-US95-16930-1
Sequence 1, Application PC/TUS9516930
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE
TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF
TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECAY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16930
FILING DATE: 27-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375,300
FILING DATE: 20-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/046W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)542-5070
TELEFAX: (617)542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid

Search completed: September 4, 2000, 00:29:28
Job time: 19938 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 19:57:05 ; Search time 299.13 seconds
(without alignments)
2859.649 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggttaattaccacgaatttg.....taaatcagattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51	1.5	4154	1 Q79037	Human SREBP-1a cod
2	47.6	1.4	5197	1 Q79038	Human SREBP-2 cod
3	43.8	1.3	3060	1 Q03098	Enod2a genomic clo
4	43.8	1.3	3856	1 Q03097	Enod2b genomic clo
5	38.2	1.1	672	1 Q70140	Cadrac cDNA. Agent
6	37.4	1.1	1981	1 T66241	Romaine lettuce vi
7	37.4	1.1	2295	1 T31994	Nonsense-mediated
8	37.4	1.1	4080	1 T31993	Nonsense-mediated
9	36.6	1.1	289	1 X07421	Borrelia burgdorfe
10	36.4	1.1	1028	1 O45606	Sequence of pATDFL
11	36.2	1.1	11871	1 X13108	Enterococcus faeca
12	36	1.1	1648	1 Q25532	Sequence of genomi
13	35.6	1.0	114955	1 X53491	Human adenosine A1
14	34.8	1.0	2758	1 T11575	Human glucagon deg
15	34.4	1.0	2344	1 X19798	Human growth regul
16	34.4	1.0	53526	1 T94101	Human PKD1 gene. H
17	34.4	1.0	53577	1 T18551	Human polycystic k
18	34.4	1.0	53577	1 T94108	Human PKD1 locus b
19	34	1.0	1193	1 X25017	Alfaifa cyclin-dep
20	33.8	1.0	924	1 V17144	Human V3 loop HIV
21	33.8	1.0	1596	1 V43619	Human secreted pro
22	33.8	1.0	1679	1 X00719	Human secreted pro
23	33.8	1.0	1863	1 X00689	Human secreted pro
24	33.8	1.0	4418	1 V74765	Staphylococcus aur
25	33.4	1.0	215	1 X12235	Human biallelic po
26	33.4	1.0	32207	1 V73805	KSHV LUR DNA (nucl
27	33.4	1.0	117213	1 V62176	HSV-2 strain SB5 C
28	33.4	1.0	137507	1 V19941	KSHV long unique c
29	33.2	1.0	2417	1 V22222	SIRE-1 cDNA from l
30	33.2	1.0	14141	1 X13331	Enterococcus faeca
31	33	1.0	6022	1 X23983	Human hgr1 DNA. A
32	33	1.0	13856	1 V74342	Staphylococcus aur
33	32.8	1.0	2277	1 V05370	Human telomerase p

34	32.8	1.0	2277	1 V13834	Homo sapiens ambig
35	32.8	1.0	2721	1 Q10212	BamHI J-I fragment
36	32.8	1.0	2721	1 Q10543	BamHI J-I fragment
37	32.8	1.0	8438	1 Q73500	DNA encoding Pseud
38	32.6	1.0	570	1 T13041	Cotton fibre-speci
39	32.6	1.0	570	1 T30258	Cotton fibre cell-
40	32.6	1.0	570	1 T70048	Cotton fibre speci
41	32.6	1.0	609	1 T62617	Cotton fibre speci
42	32.6	1.0	1476	1 T08578	Zinc finger protei
43	32.6	1.0	3680	1 T10665	Mouse bradykinin B
44	32.6	1.0	11739	1 X13087	Enterococcus faeca
45	32.4	0.9	400	1 V78220	Staphylococcus aur

ALIGNMENTS

RESULT 1
Q79037
ID Q79037 standard; DNA: 4154 BP.
AC Q79037; 1995 (first entry)
DE Human SREBP-1a coding sequence.
KW Probe; sterol regulatory element binding protein; regulatory protein;
KW cholesterol metabolism; sterol regulatory element-1; SREBP-1;
KW SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW transcription factor; low density lipoprotein; LDL; receptor;
KW 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW hypercholesterolaemia; enhancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 167..3610
FT /*tag= a
FT /product= SREBP-1a
PN WO9426922-A.
PD 24-NOV-1994.
PF 13-MAY-1994; U05300.
PR 13-MAY-1993; US-061697.
PR 01-OCT-1993; US-131365.
PA (TEXA) UNIV TEXAS SYSTEM.
PI Briggs MR, Brown MS, Goldstein JL, Wang X;
DR WPI; 95-006813/01.
DR P-PSDB; R66390.
PT New sterol regulator element binding protein - used to develop
PT prods. and screening assays for agents for reducing plasma
PT cholesterol levels (Eng)
PS Claim 31; Page 185-98; 305pp; English.
CC The sequences given in Q79037-38 encode the sterol regulatory element
CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC proteins which are involved in the regulation of genes involved in
CC cholesterol metabolism that are under the control of an associated
CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-zip)
CC transcription factors. Each have the ability to bind to SRE sequences
CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
CC found in the promoters for the low density lipoprotein (LDL) receptor
CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
CC transcription in the absence of sterols and is inactivated when sterols
CC accumulate. Human SREBP-2 contains 1141 amino acids and has 47% identity
CC with SREBP-1a. SREBP-1a was the first recognised member of this family
CC and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2
CC includes an acidic N-terminus, a highly conserved bHLH-zip motif (71%
CC identical), and an unusually long extension of 740 amino acids on the
CC C-terminal side of the bHLH-zip region. SREBP-2 possesses one feature
CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
CC production in the presence of sterols. SREBP identified in screening
CC assays, may be used to reduce plasma cholesterol levels and in
CC controlling hypercholesterolaemia and its associated diseases.
SQ Sequence 4154 BP; 714 A; 1406 C; 1280 G; 754 T;


```
Query Match      1.3%; Score 43.8; DB 1; Length 3060;
Best Local Similarity 51.8%; Pred. No. 0.0051;
Matches 99; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 633 tacatgtctccatcagcaccgaagctcaaggcccatcagaccatcatattactcacacac 692
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 1917 TACCTACTCTCTCATGAGAAACCCGCCACCAATACCAACCTCTCTCATGAGAAACCAACC 1976

Qy 693 cataaatctccaccacctcatcaccaccatcaccaccgatgccaaaaatccatgagaac 752
    ||| | ||| ||||| ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 1977 CATGAGAAATCCACCCGCGGACCACCAACCTCTCATGAGAAACCCAGACCAACCA 2036

Qy 753 cctgaacaagtggcatctccatcgattgaagatgctccagagacgaacaaactcatttg 812
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 2037 CCTCATGAGAAGCCACCACCAAGATGATGAACCACTCTCATGAGAAACCCACCACCAAGAAATAC 2096

Qy 813 gtgaacacaca 823
    ||| ||
Db 2097 CAACCACTCA 2107

RESULT 4
Q03097
ID Q03097 standard; cDNA; 3856 BP.
AC Q03097;
DE 05-JUN-1990 (first entry)
DE Enod2b genomic clone.
KW Enod2; soybean; root node; regulatory region; nodulin 75; ss.
OS Glycine max (L) Merr. cv. Williams.
FH Key Location/Qualifiers
FT misc_rna 1320..2365
FT function=regulatory region
FT /*tag= a
FT 2245..3384
FT /*tag= b
FT /*label=ORF1
FT /*notes="actual coding sequence of Enod2 encoding N-75"
FT cds 2552..3271
FT /*tag= c
FT /*label=ORF2
FT EP-349338-A.
PN 03-JAN-1990.
PD 30-JUN-1989; 306666.
PR 01-JUL-1988; US-214297.
PA (LUBR) Lubrizol Genetics.
PI Franssen H, Bisseling AH;
DR WPI; 90-009296/02.
DR P-PSDB; R04118, R04119.
DR Enod2 gene regulatory region - used for early expression of structural
PT gene in developing root node of soybean plant.
PS Disclosure; page 14-16; 18pp; English.
CC The sequences contains the Enod2a gene regulatory region (tag a) which is
CC used to construct recombinant DNA molecule for the early expression of a
CC structural gene in soybean root nodules. Two open reading frames are
CC present, ORF1 and ORF2 (R04118 and R04119 resp.) but it is thought that
CC ORF1 actually encodes the protein, N-75.
SQ Sequence 3856 BP; 1331 A; 757 C; 496 G; 1264 T; 8 Others;

Query Match      1.3%; Score 43.8; DB 1; Length 3856;
Best Local Similarity 51.8%; Pred. No. 0.0059;
Matches 99; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 633 tacatgtctccatcagcaccgaagctcaaggcccatcagaccatcatattactcacacac 692
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 2719 TACCTACTCTCTCATGAGAAACCCGCCACCAATACCAACCTCTCTCATGAGAAACCAACC 2778

Qy 693 cataaatctccaccacctcatcaccaccatcaccaccgatgccaaaaatccatgagaac 752
    ||| | ||| ||||| ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 2779 CATGAGAAATCCACCCGCGGACCACCAACCTCTCATGAGAAACCCAGACCAACCA 2838

Qy 753 cctgaacaagtggcatctccatcgattgaagatgctccagagacgaacaaactcatttg 812
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
```

```
Db 2839 CCTCATGAGAAGCCACCACCAGATGATGAACCACTCTATGAGAAACCAACCACCAAGAAATAC 2898

Qy 813 gtgaacacaca 823
    ||| ||
Db 2899 CAACCACTCA 2909

RESULT 5
Q70140/c
ID Q70140 standard; cDNA; 672 BP.
AC Q70140;
DE 23-FEB-1995 (first entry)
DE Cadrac cDNA.
KW Cadrac; Cadracpro; Cadop; antisense; Parkinson disease; diagnosis;
KW therapy; APP gene; amyloid precursor protein; dopamine receptor;
KW calcium channel; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..672
FT /*tag= a
PN WO9415967-A.
PD 21-JUL-1994.
PF 30-DEC-1993; E03722.
PR 31-DEC-1992; CA-086576.
PR 06-OCT-1993; US-132169.
PA (BERG/) PREDDIE R E.
PA (BERG/) BERGMANN J F.
PI Bergmann JE, Preddie RE;
DR WPI; 94-249143/30.
DR P-PSDB; R57348.
DR Agents for prevention and treatment of Parkinson's Disease -
PT comprises cadrac gene anti-sense to regions of human APP gene
PS Disclosure; Page 35; 47pp; English.
CC Cadrac cDNA encodes a protein, Cadop, with potential to be a common
CC domain for interacting with the human D-2 dopamine receptor and the
CC human brain Ca2+ channel protein. Cadrac comprises nucleotides
CC 553-1224 of the antisense strand of the human amyloid precursor
CC protein mRNA. Cadrac and Cadop are used for the diagnosis and
CC therapy of Parkinsonism.
SQ Sequence 672 BP; 116 A; 210 C; 139 G; 207 T;

Query Match      1.1%; Score 38.2; DB 1; Length 672;
Best Local Similarity 61.6%; Pred. No. 0.1;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 959 ggagaagcgcggaagattgtgattccaggagacagcggaaggggatgaagatgaggatga 1018
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 394 GGAGCAGACACAGACTATGCAGATGCTGGAAGAAGAGCCGATGATGACGAGGACGA 335

Qy 1019 tgaagatagatccgggagactatgtctcagggaac 1057
    ||| ||||| ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 334 TGAGGATGGTGTAGGAGTAGAGGAGGCGCTGAGGAACC 295

RESULT 6
T66241
ID T66241 standard; cDNA; 1981 BP.
AC T66241;
DE 28-JUL-1997 (first entry)
DE Romaine lettuce violaxanthin de-epoxidase cDNA.
KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; lettuce; ss.
OS Lactuca sativa L. cv. Romaine.
FH Key Location/Qualifiers
FT misc_difference 26..29
FT /*tag= a
FT /*note= "bases 26-29 are illegible in Fig 1"
FT misc_difference 66..72
FT /*tag= b
FT /*note= "bases 66-72 are illegible in Fig 1"
FT misc_difference 105..110
```

```

FT FT misc_difference 147..149 /tag= c
FT FT /note= "bases 105-110 are illegible in Fig 1"
FT FT /tag= d
FT FT /note= "bases 147-149 are illegible in Fig 1"
FT FT misc_difference 186..189 /tag= e
FT FT /note= "bases 186-189 are illegible in Fig 1"
FT FT misc_difference 226..227 /tag= f
FT FT /note= "bases 226-227 are illegible in Fig 1"
FT FT cds 235..1656 /tag= g
FT FT transit_peptide 235..609 /tag= h
FT FT mat_peptide 610..1653 /tag= i
FT FT WO9717447-A2.
PN PD 15-NOV-1996; U18291.
PF PD 07-NOV-1996; U18291.
PR PD 07-NOV-1995; US-006315.
PR PD 06-AUG-1996; US-023502.
PA (CALJ) CALGENE INC.
PI Budos RC, Rockholm DC, Yamamoto HY;
DR WPI; 97-281036/25.
DR P-PSDB; W09874.
PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
PT sensitivity of a plant to light
PS Claim 3; Fig 1; 41pp; English.
CC A cDNA clone (T66241) codes for the 55 kDa violaxanthin de-epoxidase
CC (VDE) (W09874) of romaine lettuce. VDE was purified from romaine
CC lettuce chloroplasts and 2 tryptic peptides were used to develop
CC primers (see also T66244-45), which amplified a partial VDE
CC sequence. The amplified sequence was then used to screen a lettuce
CC cDNA library, and the 1981 bp DNA sequence was identified. VDE
CC nucleic acids (see also T66242-43), in sense or antisense
CC orientation, can be used in genetic constructs, pref. also contg. a
CC plastid translocation sequence, to modify VDE levels in plants.
CC Increased levels result in the plant being tolerant of increased
CC light and therefore more productive and/or more resistant to
CC disease. Underexpression of VDE increases photosynthetic
CC efficiency under low light. The photosensitivity of a range of
CC crops, trees and ornamentals can be modified.
SQ Sequence 1981 BP; 608 A; 337 C; 433 G; 577 T;

Query Match 1.1%; Score 37.4; DB 1; Length 1981;
Best Local Similarity 62.1%; Pred. No. 0.38;
Matches 59; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 960 gagaagcgcggaagtgtgattcaggagacgcggaagggatgaagatgagatgat 1019
Dn 1405 GAGAAGTTGTTGATTAAGAAGCTGTAGAGATAGAGAGAGAGGTTGAAAAGAGGTGCAG 1464

QY 1020 gaggatagtgatccggggagagactatgtctcagg 1054
Dn 1465 AAGGTTAGAGATAGTACGATGACTTGTGTTTCAGAG 1499

RESULT 7
T31994
ID T31994 standard; DNA; 2295 BP.
AC T31994;
DE 24-NOV-1996 (first entry)
KW Nonsense-mediated mRNA decay 2 gene C-terminal (1089-3383).
OS Nonsense-mediated mRNA decay; NMD2; Upflp; inhibition; ss.
PN Saccharomyces cerevisiae.
PS WO9622301-A1.
PD 25-JUL-1996.
PF 27-DEC-1995; U16930.
PR 20-JAN-1995; US-375300.
PI (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PT Isolated nonsense-mediated mRNA decay gene and protein - used to
PT develop methods for inhibiting the decay pathway for producing
PT heterologous or endogenous proteins
CC Claim 1; Page 47-49; 73pp; English.
CC The NMD2 gene (T31993) is named after its role in the Nonsense-
mediated mRNA Decay pathway. The protein, Nmd2p, binds to Upflp.

```

```

DR WPI; 96-354469/35.
DR P-PSDB; W01897.
PT Isolated nonsense-mediated mRNA decay gene and protein - used to
PT develop methods for inhibiting the decay pathway for producing
PT heterologous or endogenous proteins
PS Claim 2; Page 52-54; 73pp; English.
CC The NMD2 gene (T31993) is named after its role in the Nonsense-
CC mediated mRNA Decay pathway. The protein, Nmd2p, binds to Upflp.
CC A C-terminal fragment of the protein (T31994) also binds Upflp and,
CC when overexpressed in the host cell, the fragment inhibits the
CC function of Upflp, thereby inhibiting the nonsense-mediated
CC mRNA decay pathway.
SQ Sequence 2295 BP; 891 A; 359 C; 447 G; 598 T;

Query Match 1.1%; Score 37.4; DB 1; Length 2295;
Best Local Similarity 53.0%; Pred. No. 0.41;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 889 gtgaagtgaacggttaaaagaataaaatcaggagcatgttcagcgcaatggccat 948
Dn 1589 GTGTGATTTACTAGGAGAAGATGAAGACGCGGAGATTAAGTACACGACAGAGTCAG 1648

QY 949 cgaggagtaaggagagcgcggaagattgtgattcaggagacgcggaagggatgaag 1008
Dn 1649 CCCCAGGAAAACATCAGGCAAGCAAGCAAGTGAAGATGAAGACGATGAGGACGATG 1708

QY 1009 atgaggatgagatgagatgattcggggga 1039
Dn 1709 ACGAGGATGATGACGATGACGATGACGATGA 1739

RESULT 8
T31993
ID T31993 standard; DNA; 4080 BP.
AC T31993;
DE 24-NOV-1996 (first entry)
KW Nonsense-mediated mRNA decay 2 gene.
OS Nonsense-mediated mRNA decay; NMD2; Upflp; inhibition; ss.
PN Saccharomyces cerevisiae.
PS Key Location/Qualifiers
FT cds 362..3633 /tag= a
FT intron 368..480 /tag= b
FT misc_feature 368..373 /tag= c
FT /note= "5' splice site"
FT misc_feature 426..432 /note= d
FT /note= "branchpoint"
FT misc_feature 478..480 /tag= e
FT /note= "3' splice site"
FT tata_signal 143..149 /tag= f
FT protein_bind 164..177 /tag= g
FT /label= Upflp_binding_consensus

PN WO9622301-A1.
PD 25-JUL-1996.
PF 27-DEC-1995; U16930.
PR 20-JAN-1995; US-375300.
PI (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PT He F, Jacobson AS;
DR WPI; 96-354469/35.
DR P-PSDB; W04896.
PT Isolated nonsense-mediated mRNA decay gene and protein - used to
PT develop methods for inhibiting the decay pathway for producing
PT heterologous or endogenous proteins
PS Claim 1; Page 47-49; 73pp; English.
CC The NMD2 gene (T31993) is named after its role in the Nonsense-
mediated mRNA Decay pathway. The protein, Nmd2p, binds to Upflp.

```

```
CC A C-terminal fragment of the protein (T31994) also binds Upflp and,
CC when overexpressed in the host cell, the fragment inhibits the
CC function of Upflp, thereby inhibiting the nonsense-mediated
CC mRNA decay pathway.
SQ Sequence 4080 BP; 1449 A; 646 C; 800 G; 1185 T;

Query Match
Best Local Similarity 1.1%; Score 37.4; DB 1; Length 4080;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 889 gtsaagttagcaggttaaaagaataaaataccaggagcatgttcagcgacgaatggcccat 948
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3038 GTCTCGATTACTAGGAGAAGATCAAGACGCGGAGATAAAGTACACCGACACAGAGTCAG 3097

QY 949 cgagaggaaggaagggcggaagattgtgtattcaggagacagcggaaggagatgaag 1008
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3098 CGCAGGAANACATCAGCGAAGCAACAGCAAGAAAGTGAAGATGAAGACGATGAGGACGATG 3157

QY 1009 atgagatgatgaggaatgattccgggga 1039
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3158 ACGAGGATGATGACGATGACGATGACGATGA 3188

RESULT 9
X07421
ID X07421 standard; DNA; 289 BP.
AC X07421;
DT 08-JUN-1999 (first entry)
DE Borrelia burgdorferi surface antigen p39.5 clone 12-1 3' end.
KW Lyme disease; surface antigen; p39.5; diagnosis; prevention;
KW vaccine; antisense; therapy; treatment; primer; probe;
OS antibody; DNA; ds.
OS Borrelia burgdorferi.
PN W09900413-A1.
PD 07-JAN-1999.
PF 29-JUN-1998; 013551.
PA 30-JUN-1997; US-051271.
PA (TULA ) TULANE EDUCATIONAL FUND.
PI Philipp MT.
PI WPI: 99-095676/08.
DR New nucleic acid encoding the p39.5 antigen of Borrelia burgdorferi
PT - and related vectors, transformants, antibodies and polypeptides,
PT for diagnosis, prevention and treatment of Lyme disease
PS Claim 3; Page 59; 89pp; English.
CC The sequence is that of Borrelia burgdorferi surface antigen p39.5
CC clone 12-1 (3'end). It can be used in the production of p39.5 or
CC fragments of it which may be used to raise antibodies to, and in the
CC development of vaccines against Lyme disease. The sequence can
CC also be used for making primers and probes for diagnosis, also in
CC DNA vaccines, as antisense therapeutics and for drug screening.
CC Antibodies can be used as diagnostic (immunoassay) reagents, for
CC treating Lyme disease, for affinity purification, for drug
CC screening and to produce anti-idiotypic antibodies (used in the
CC same way as p39.5 to induce an immune response).
SQ Sequence 289 BP; 83 A; 29 C; 104 G; 73 T;

Query Match
Best Local Similarity 1.1%; Score 36.6; DB 1; Length 289;
Matches 72; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 914 aaaaatcaggagcatgttcagcgacgaatggccatcgaggagtaaggaagcgcgcaa 973
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 AACAGCAAAAGTGGTGATGCCGAAGAGTGTAAATGGGATGCTAAGGGGATAAAGGGGAT 170

QY 974 gattgttcaggagacacgcggaaggggatgaagatgagatgagtagtgatttc 1033
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 TGTTGATGCTGCTGACGAAGCTCATCCGAAGGAGGAACTGCGATGTCGTCGGTGCATCC 230

QY 1034 cggggagacta 1044
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 TGCTGAACACTA 241
```

```
RESULT 10
Q45606
ID Q45606 standard; DNA; 1028 BP.
AC Q45606;
DT 04-DEC-1993 (first entry)
DE Sequence of PATDFLAG encoding signal peptide, H4VL, linker and
DE FLAG adapter.
DE Single chain antibody; SCFV1: human subgroup 4 germline antibody;
KW variable light; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT 293..775
FT /*tag= a
FT /note= "ENCODES SIGNAL, H4VL, LINKER
FT see R38320 FT"
FT 784..816
FT /*tag= c
FT /product= flag peptide
PN W09312231-A.
PD 24-JUN-1993.
PF 13-DEC-1991; AU0583.
PR 13-DEC-1991; WO-AU0583.
PR (DOMC ) DOW CHEM AUSTRALIA LTD.
PI Johnson KS, Mezes PS, Richard RA;
PI WPI: 93-214173/26.
DR P-PSDB; R38320, R40753.
DR New composite antibody binding to tumour associated TAG-72
PT antigen - includes light chain variable region from human
PT subgroup 4 germline gene, useful, opt. as conjugate, for
PT diagnosis or treatment of cancer
PS Example: Figure 28; 150pp; English.
CC The plasmid PATGFLAG was generated from pSCFVUHH (see Q45605)
CC to incorporate a flag-coating sequence 3' to any human VH gene
CC to be expressed contiguously with Hum4 VL. The plasmid PATGFLAG,
CC when digested with XhoI and Nhe I and purified becomes the human
CC CH discovery plasmid contg. Hum4 VL in this SCFV format. pSC49FLAG
CC contains murine antibody CC49VH inserted into sites of XhoI - NheI
CC of PATDFLAG. It was evaluated for biological activity with the
CC purpose of serving as a positive control for the FLAG assay
CC system to detect binding to TAG-72.
SQ Sequence 1028 BP; 289 A; 226 C; 244 G; 269 T;

Query Match
Best Local Similarity 1.1%; Score 36.4; DB 1; Length 1028;
Matches 93; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 969 gcaagattgtgattcagagacagcggaaggagatgaagatgagatgagatgag 1028
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
704 CGCGACGATCGGAAAAGGATGTCGCCAAGAGATGACGCTAAGAAAGACGATGCTAAA 763

QY 1029 gattcggggagactatgtctcagggaactactattattgt-tcgaagaccacaaaccga 1087
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
764 AAGGACCTCGAGACAATGTCGCTAGCGACTACAAGGACGATGATGACAAATAAAACCTA 823

QY 1088 gqdtctacgggcacacaactctctcgaagaagtagatagatgctcaataa 1137
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GCGATCAATCCGTCAAAACATCATCTTACATAAAGTCACCTTGGTGATCAA 873

RESULT 11
X13108/c
ID X13108 standard; DNA; 11871 BP.
AC X13108;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:171.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
```


WPI: 99-229400/19.
New antisense oligonucleotides used in treatment of, e.g. pulmonary vasoconstriction
Disclosure: Page 37: 120pp; English.
The specification describes antisense oligonucleotides (X52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences X5271-74. These multiple target oligonucleotides (specifically X55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.
Sequence 114955 BP: 6071 A; 29417 C; 36712 G; 36721328 T.

Query Match	1.08;	Score 35.6;	DB 1;	Length 114955;
Best Local Similarity	38.0%;	Pred. No. 20;		
Matches	89;	Conservative 34;	Mismatches 111;	Indels 0; Gaps 0;
QY 687	caacaccatcaatctccaccacctcatccaccatcaccaccgcatgccaaaaatccat	746		
Db 79264	CVGCVCAAGNNHHNNNSCCVGGCCGCGVCVCGAVVAGVCAACCAACCAVVGCVGAACVGA 79205			
QY 747	gagaaacctgaacaagtggcatctctcatcgattgaagatgtctccagagacgaaacaaact	806		
Db 79204	GAGCACAVGCAAAVAVVACAVGCCACAGAVGGGAVVVGAVGVCVAAAIVVVCACV 79145			
QY 807	catttggttgaaccacaaagtccaaaaagcccgacgaatatgaaagaggagctctctcg	866		
Db 79144	GGCAVGGCCACVCGAGVAGVAVVACGCCCCVGVNNHHNNNSCCTGCCCGTGCTTCAG 79085			
QY 867	ttactagttaacatgtctccgagtgaagtgaaacgggttaagaataaaaaatca	920		
Db 79084	TTAGATCAAAACATTGCTGAAACCTGAAGAGACGACATGCAAAATATTACAGATCCCA 79031			

RESULT 14
Tl1575/c
ID Tl1575 standard; cDNA; 2758 BP.
AC Tl1575;
DT 09-AUG-1996 (first entry)
DE Human glucagon degrading enzyme gene.
KW Glucagon degrading enzyme; catalyst; cleavage; selectin; human; primer;
KW vasoactive intestinal peptide; VIP; pancreatic carcinoma cell line; PCR
KW amplification; polymerase chain reaction; probe; expression vector;
KW eukaryote; SV40 promoter; COS-7; ss.
OS Homo sapiens.

PH	Key	Location/Qualifiers
FT	FT	100..2694
FT		/*tag= a
FT		/product= glucagon degrading enzyme

PN J08023972-A.
PD
PD 30-JAN-1996.
PF 19-JUL-1994; 187936.
PF 19-JUL-1994; JP-187936.
PR (SUNR) SUNTORY LTD.
PA WPI: 96-133414/14.
DR P-PSDB; R93021.
DR
PT New glucagon decomposing enzyme, and DNA encoding it - for specifically cleaving glucagon and vasoactive intestinal peptide, in the prevention and treatment of diseases caused by excess glucagon

PT and VIP
PS Claim 4; Page 3; 18pp; Japanese.
CC This is the nucleotide sequence of a novel isolated gene encoding a
CC glucagon degrading enzyme (GDE) of mol. wt. 83 kD. The enzyme has a
CC pH optimum of 6.8 and catalyses the cleavage of glucagon, vasoactive
CC intestinal peptide and selectin (R93022-4). The gene was isolated
CC from a human pancreatic carcinoma cell line HPC-Yo cDNA library by
CC screening the library with an anti-GDE peptide antibody, amplifying
CC the inserts with the primers T1803-4 and probing the fragments with
CC the probe T18905. This screening resulted in the full length clone
CC designated lambda GDE4-2. The coding region of the clone was
CC subsequently PCR amplified by the primers T11576-7 and inserted into
CC the eukaryotic expression vector pKOCF under control of the SV40
CC promoter for production of the protein in COS-7 cells. The protein
CC is useful in preventing and treating diseases characterised by an
CC excess of glucagon or vasoactive intestinal peptide.
SQ Sequence 2758 BP; 968 A; 416 C; 483 G; 891 T

Query Match	1.0%	Score 34.8;	DB 1;	Length 2758;
Best Local Similarity	65.4%	Pred. No. 3;		
Matches 51;	Conservative	0;	Mismatches 27;	Indels 0;
Gaps 0;				

Qy 1550 tatgtctgggtggttcaacatcttaatgatcatatatatgttggtcacaactgcgtatccat 1709
| | | | | | | | | | | | | | | | | | | | |
Db 696 TCTATCTTCTTAATTCACACTCTTTATTGATATAAGTCTACATCAAACTCATGATTAAAT 637

OY	1710	ggtgacctgttcagac	1727
Db	636	TGTCACGCCGTTTAAAC	619

RESULT 15
X19798
ID X19798 standard; cDNA; 2344 BP.

AC	X19798:	Human growth regulator protein GRREG encoding CDNA
DT	09-JUN-1999	(first entry)
DE		Human growth regulator protein GRREG encoding CDNA
KW		Human; growth regulator protein; GRREG; cancer; ss.
OS		Homo sapiens.

EH	Key	Location/Qualifiers
FT	CDS	226. .2250
ET		/*tag= a

PN WO9902680-A1.

PD 21-JAN-1999.

PF 30-JUN-1998; U13409

PR 11-JUL-1997; US-893852.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Corley NC, Lal P, Shah P;

DR WPI; 99-120881/10.

Novel human growth regulator protein - useful in the treatment of cancers

PS Claim 7; Fig 1; 70pp; English.

CC The present sequence encodes a human growth regulator protein,

CC designated GRREG. The protein can be used in methods to treat cancer.
CC The protein can also be used in binding assays to detect (antagonists
CC of GRREG. A fragment from the nucleic acid sequence encoding GRREG can
CC be used as a probe for detecting GRREG encoding sequences (especially
CC in PCR amplified samples).

Sequence	2344 BP;	556 A;	667 C;	709 G;	412 T;
SQ					

Query Match	1.0%;	Score 34.4;	DB 1;	Length 2344;
Best Local Similarity	78.8%;	Pred. No. 3.6;		

985 aggaagacagccgaaaggggattcaagattcaagattcaagattcaattccgg 1036

db 1268 AGGACACAGAGGAGAGGAGAGATGAGGAGAGATGAGGAGAGTCACTCTGG 1319

Search completed: September 4, 2000, 00:38:25
Job time: 16880 sec

21

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:24:33 ; Search time 4457.97 Seconds
(without alignments)
1949.715 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgagattaatgctg.....aaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: gb_est1.*
21: gb_est2.*
22: gb_est3.*
23: gb_est4.*
24: gb_est5.*
25: gb_est6.*
26: gb_est7.*
27: gb_est8.*
28: gb_est9.*
29: gb_est10.*
30: gb_est11.*
31: gb_est12.*
32: gb_est13.*
33: gb_est14.*
34: gb_est15.*
35: gb_est16.*
36: gb_est17.*
37: gb_est18.*
38: gb_est19.*
39: gb_est20.*
40: gb_est21.*
41: gb_est22.*
42: gb_est23.*
43: gb_est24.*
44: gb_est25.*
45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: em_est20.*
53: em_est21.*
54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*
59: gb_est33.*
60: gb_est34.*
61: gb_est35.*
62: gb_est36.*
63: gb_est37.*
64: gb_est38.*
65: em_est27.*
66: em_est28.*
67: em_est29.*
68: em_est30.*
69: gb_est39.*
70: gb_est40.*
71: gb_est41.*
72: gb_est42.*
73: gb_est43.*
74: gb_est44.*
75: em_est31.*
76: em_est32.*
77: em_est33.*
78: em_est34.*
79: gb_est45.*
80: gb_est46.*
81: gb_est47.*
82: em_est35.*
83: em_est36.*
84: em_est37.*
85: gb_est48.*
86: gb_est49.*
87: gb_est50.*
88: gb_est51.*
89: gb_est52.*
90: gb_est53.*
91: gb_est54.*
92: gb_est55.*
93: gb_gss1.*
94: gb_gss2.*
95: gb_gss3.*
96: gb_gss4.*
97: em_gss1.*
98: em_gss2.*
99: em_gss3.*
100: em_gss4.*
101: gb_gss5.*
102: gb_gss6.*
103: gb_gss7.*
104: gb_gss8.*
105: gb_gss9.*
106: em_gss5.*
107: em_gss6.*
108: em_gss7.*
109: em_gss8.*
110: em_gss9.*
111: em_gss10.*
112: em_gss11.*
113: gb_gss10.*
114: gb_gss11.*
115: em_gss12.*
116: gb_gss12.*

117: gb_gss13:*
 118: gb_gss14:*
 119: gb_gss15:*
 120: gb_gss16:*
 121: gb_gss17:*
 122: gb_gss18:*
 123: gb_gss19:*
 124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	745.4	37.8	788	25	AA391707	AA391707 LD11632.5
2	590.4	30.0	592	41	AI546088	AI546088 LD39580.5
3	564.2	28.6	578	26	AA439767	AA439767 LD14421.5
4	557.6	28.3	584	33	AA990888	AA990888 LD34294.5
5	50.6	2.6	391	45	AI852939	AI852939 UI-M-BH0-
c 6	46.8	2.4	1101	122	CNS00293	AL097377 Drosophil
7	46.4	2.4	928	122	CNS000DKY	AL071865 Drosophil
8	46.2	2.3	282	81	C94400	C94400 C94400 Dict
c 9	45.8	2.3	946	122	CNS008EV	AL052494 Drosophil
10	45.4	2.3	455	80	C24663	C24663 C24663 Dict
c 11	45.2	2.3	928	122	CNS00DKY	AL071865 Drosophil
12	44.6	2.3	311	48	AU071134	AU071134 AU071134
c 13	44.6	2.3	825	122	CNS00BTL	AL058023 Drosophil
14	44.4	2.3	315	31	AA839308	AA839308 vo40d03.r
15	43.8	2.2	202	70	AW307211	AW307211 sf54b11.y
16	43.8	2.2	360	80	C11831	C11831 C11831 Yuj1
17	43.8	2.2	437	80	AW702259	AW702259 TgESTz74
c 18	43.8	2.2	995	122	CNS003IU	AL064551 Drosophil
c 19	43.6	2.2	427	45	AI874339	AI874339 tz64e05.x
c 20	43.4	2.2	240	48	AU073211	AU073211 AU073211
21	43.2	2.2	1201	123	CNS016A0	AL106482 Drosophil
c 22	42.6	2.2	522	42	AI673293	AI673293 tw59g09.x
c 23	42.2	2.1	348	40	AI496844	AI496844 fb64e06.x
24	42.2	2.1	563	48	AU038220	AU038220 AU038220
25	42.2	2.1	726	39	AI406223	AI406223 GH26449.5
c 26	42.2	2.1	213	37	AI251434	AI251434 qu74e10.x
27	42.2	2.1	264	81	C90816	C90816 C90816 Dict
c 28	42.2	2.1	312	46	AI928293	AI928293 wp12a09.x
29	42.2	2.1	517	47	AU033605	AU033605 AU033605
c 30	42.2	2.1	786	122	CNS00609	AL065615 Drosophil
c 31	41.8	2.1	197	41	AI559457	AI559457 tq44b03.x
32	41.8	2.1	326	23	AA242229	AA242229 mw26f02.r
33	41.6	2.1	245	70	AW311640	AW311640 sg43d06.y
c 34	41.4	2.1	507	64	AW165325	AW165325 fd99f02.y
35	41.4	2.1	690	81	C93837	C93837 C93837 Dict
36	41.4	2.1	1101	122	CNS003BD	AL064091 Drosophil
37	41.2	2.1	200	81	C93770	C93770 C93770 Dict
c 38	41.2	2.1	312	31	AA780779	AA780779 ag14b08.s
39	41.2	2.1	314	63	AW101789	AW101789 sd70d10.y
c 40	41.2	2.1	351	23	AA253430	AA253430 zr77f02.s
c 41	41.2	2.1	761	44	AI812108	AI812108 tw77d04.x
42	41.2	2.1	986	123	CNS01678	AL106382 Drosophil
c 43	41.2	2.1	208	39	AI416651	AI416651 sa10a03.x
44	41.2	2.1	301	47	AI121328	AI121328 DKEzp762P
45	41.2	2.1	467	31	AA801779	AA801779 GM12988.5

ALIGNMENTS

RESULT 1
 AA391707 788 bp mRNA EST 28-NOV-1998
 LOCUS AA391707 Drosophila melanogaster embryo Bluescript
 DEFINITION LD11632.5prime LD Drosophila melanogaster cDNA clone LD11632 5prime, mRNA sequence.
 Drosophila melanogaster cDNA clone LD11632 5prime, mRNA sequence.

AA391707
 AA391707.1 GI:2044683
 EST.
 fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 788)
 Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
 Brokstein,P., Lewis,S. and Rubin,G.M.
 BDGP/HMI Drosophila EST Project
 JOURNAL
 Unpublished (1997)
 COMMENT
 On Jan 14, 1998 this sequence version replaced gi:1797339.
 Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 LSA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Plate: 116 row: C column: 8
 High quality sequence stop: 694.
 Location/Qualifiers
 1..788
 /organism="Drosophila melanogaster"
 /db_xref="BDGP_EST:BDcln010843"
 /db_xref="taxon:7227"
 /clone_lib="LD11632"
 /clone_lib="LD Drosophila melanogaster embryo Bluescript"
 /sex="male and female"
 /dev_stage="0 to 24 hours mixed stage embryonic"
 /lab_host="SOLR"
 /note="Organ: embryo; Vector: Bluescript SK; Site:1:
 EcoRT; Site:2: XhoI; Constructed using Stratagene ZAP-cDNA
 Synthesis kit. Oligo dT-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"
 BASE COUNT 175 a 223 c 190 g 200 t
 ORIGIN
 Query Match 37.8%; Score 745.4; DB 25; Length 788;
 Best Local Similarity 98.9%; Pred. No. 1.4e-202;
 Matches 782; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 QY 19 tgatttctggtctggtactcacagcattgctggtataagagtcggagcagagagagtaa 78
 Db 1 TGATTTCGTGGTGGACTACACAGCATTGC-GGTATAAGAGTTCGGGACAGAGAGTAA 59
 QY 79 gatttgggaaggaatcccgctcggtaggagtagcttagcttagcagagtagcagca 138
 Db 60 GATTTTCGGGAAGGAATCCCGTCCGGTA-GGACTACTAGCATTCGCAAGTGACGTCCAGCA 118
 QY 139 accggaggacccccactataatccgcatcacaccatccttaaccacaaaccaatgac 198
 Db 119 ACCGGA-GACCCCCAACTAGTAATCCGATCCCATCCATTCATCCCAACAAATGAC 177
 QY 199 atcttgagacctcaccagcagcattgcttctggtgttcttcttagtactggtcgtcttt 258
 Db 178 ATCTTGAGACCTCACCAGCAGCATTCCTTCGTGTTCTTCATAGTACTGGCATCGCTTT 237
 QY 259 atggogtcttcttacttttcttcgacgcttcttcaagagtgatgcatgactaccctagatg 318
 Db 238 ATGGCGCTTCTTACTTTTTCGACCGCTTCTTCAAGAGTTCGATGCTACTACCCGTACGATG 297
 QY 319 ccttctcagaacacacgggctgagtgatataatttcactgacctccactggcacacagtg 378
 Db 298 CCTTCTCAAGAACACCGGCTGGAGTATTAATTTATGAGCCCTCCACTGGCACACGAGTG 357
 QY 379 cctttaacaggaccctcctacgtggggatctgcccgttaacagctgcacccggagagtaa 438
 Db 358 CCTTTAACAGGACCCCTCCTACGCTGGGATCTCCCGTAACAGCTGCACCCGGAGAGTAA 417
 QY 439 tgatcacccagctttaatgtaggagtcctggtcacctttctctgtcccgatcggtctga 498

Db	418	TGATCACAGCTTTAATGTAGAGTCCGGTCACCTTTTCTGTGCTCCGATCGGTCTGA	477
Qy	499	tctgtctatgacctactcttcagcagtggtgaaacaagacagctcttcgctgtatcatc	558
Db	478	TCCTGCTCATTTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTTCCTGCTGTATCCT	537
Qy	559	cgcctgtggagtcctctgacgtggaattctactgcccggcgctcaactgcccgttgg	618
Db	538	CGCCCGTTGGATGTCTCTGCAGCTGGAAATCTACTGCTGGCTCAACTTGCCGTTGG	597
Qy	619	aggagatcggaatactacatcacacccttgcctctgctctgttggtgcacgagatggac	678
Db	598	AGGAGATCGGATACTACATCACACCCCTTGCTGCTGCTTGGTGGTGACGAGATGGAC	657
Qy	679	acgcctggccctgtgatggagatgtgcctgtcacccgggttttggaataaagttcatct	738
Db	658	ACGCCCTGGCCGCTGTGATGAGGATGTGCTGTACCCGGGTTGGATAAAGTTCATCT	717
Qy	739	tctgcctgccttagcatcacagcgagctctcccacgaccacttaaacagtctacgttgg	798
Db	718	JCTGCCCTGCCGTTAGCATACACGGAGCTCTCCACGACCACTTAAACAGTCTACGGTGG	777
Qy	799	tccgcaagcta	809
Db	778	TCCGCAAGCTA	788
RESULT	2		
LOCUS	A1546088	592 bp	22-MAR-1999
DEFINITION	LD39580.5prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD39580 5prime, mRNA sequence.		
ACCESSION	A1546088		
VERSION	A1546088.1	GI:4463461	
KEYWORDS	EST.		
SOURCE	fruit fly,		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.		
TITLE	BGP/HHMI Drosophila EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Harvey, D. G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 USA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu hit genomic sequence AC005465; hit genomic sequence AC005466 Plate: 395 row: G column: 8 High quality sequence stop: 511. Location/Qualifiers 1. 592 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="LD39580" /clone_lib="LD Drosophila melanogaster embryo pOT2" /sex="male and female" /dev_stage="0 to 24 hours mixed stage embryonic" /lab_host="XLI Blue" /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."		
FEATURES	source		
BASE COUNT	142 a	150 c	159 g
ORIGIN			141 t
Query Match	30.08;	Score 590.4;	DB 41;
Best Local Similarity	99.8%;	Pred. No. 3.2e-158;	Length 592;

/db_xref="taxon:7227"
/clone="LD14421"
/clone_lib="LD Drosophila melanogaster embryo Bluescript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: Bluescript SK; Site_1:
EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dT-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT 134 a 163 c 135 g 146 t
ORIGIN

Query Match 28.6%; Score 564.2; DB 26; Length 578;
Best Local Similarity 98.6%; Pred. No. 1e-150;
Matches 569; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 11 attaatgctgatttctgctgactacacagcattgctggtataagagtgaggaccag 70
Db 2 ATAAATGGCGGATTTCTGGTCTGACTACACAGCATTTGCTGTATAGGAGTCCGGACCCAG 61
Qy 71 agagtaagatttcgggaaggaatcccgtaggagtagctagcttagcttagcttagc 130
Db 62 AGGAGTAAGATTTCGGGAAGGAATCCCGTAGGACTACTAGCATTCGCAAGTGAC 121
Qy 131 gtccagcagcggagagaccacacacacacacacacacacacacacacacacacac 190
Db 122 GTCCAGCAACCGAGAGACCCCACTGTAGATTCGGATCCGATCACCATCCTTAATCCCAACAAA 181
Qy 191 ccaatgacatttggagactcaccagcagcattgcttctgcttcttcttcttcttct 250
Db 182 CCAATGACATCTTGAGACCTCACCAGCATGATCCCTTCGTTCTTCTATAGTACTGCG 241
Qy 251 atcgcttatggcgttcttcttcttcttcttcttcttcttcttcttcttcttctt 310
Db 242 ATCGCTTTATGGCGTCTTTACTTTTTCACCGCTTCTTCAAGAGTTGCGATGACATGCC 301
Qy 311 gtacgactcttctcaagaacacggcgtgagtagataatttcatgagcctccactgca 370
Db 302 GTACGATGCTTCTCAAGAACACCGGCTGAGTATAAATTTTCATGAGCTCCACTGGCA 361
Qy 371 cagcagtgctttaaagagacccctcactgctgggagctctgctggttaacagctgacccg 430
Db 362 CACGAGTGCTCTTAACAGGACCCCTCTAGCTGGGGATCTGCCGTAAACAGTGCACCCG 421
Qy 431 gagagtaagatcaccagctttaaagtaggagctcctgctcacttcttcttcttctt 490
Db 422 GAGAGTAATGATCACAGAGCTTTAAATGTAGGAGTCTGCTGCTACCTTTTCTGCTCCCGAT 481
Qy 491 cggctgactctgctcattgcccactatcttcaagcagtggtgagacagcagctcttctg 550
Db 482 CGGTCTGATCTGCTCATGCGCACTATCTTACAGAGTGGTGAACAGACAGCTCTTCGTC 541
Qy 551 tgtatctcgcgcttgagtgcttccctgtgctgagctggaa 587
Db 542 TGTATACTCGCCCGTGGAGTCCCTGTGCAGCTGGGA 578

RESULT 4
AA990888 584 bp mRNA EST 24-NOV-1998
LOCUS LD34294.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
DEFINITION melanogaster cDNA clone LD34294 5prime, mRNA sequence.
ACCESSION AA990888
VERSION AA990888
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 584)

AUTHORS
TITLE
JOURNAL
COMMENT

Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
On Apr 7, 1998 this sequence version replaced gi:3036104.
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 342 row: H column: 10
High quality sequence stop: 583.

FEATURES
source

Location/Qualifiers
1..584
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD34294"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
BASE COUNT 137 a 164 c 130 g 153 t
ORIGIN

Query Match 28.3%; Score 557.6; DB 33; Length 584;
Best Local Similarity 98.3%; Pred. No. 7.9e-149;
Matches 574; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 11 attaatgctgatttctgctgactacacagcattgctggtataagagtgaggaccag 70
Db 1 ATTAATGCTGATTTCTGGTCTGGACTACACAGCATTTGCTGTATAGGAGTCCGACCCAG 60
Qy 71 agagtaagatttctggaaggaatcccgctccggt-aggagactactgattcgcgaagta 129
Db 61 AAGAGTAGATTTCGGGAGAGATCCCGTCCGGTAAGGGATTATTAGATTTCGCAAGTGA 120
Qy 130 cgtccagcaacggaggaccccaactgttagaatccgcataccatccttaatacccaaaa 189
Db 121 GTCCAGCAACCGGAGGACCCCAACTGTAGAAATCCGATCACCATCTTAATCCCAACAA 180
Qy 190 accaatgacatttagac 249
Db 181 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTCTTCATAGTACTGG 240
Qy 250 catcgctttatggcgttcttcttcttcttcttcttcttcttcttcttcttcttctt 309
Db 241 CATCGCTTTATGGCGTCTTTTACTTTTTCGACCGCTTCTTCAAGAGTTGCATGCATACC 300
Qy 310 cgtacgctgcttctcctcaagaacacccggctgagtagataaaattctatgagctccactg 369
Db 301 CGTAGATGCCTTCTCCTCAAGAACACCGGCTGAGTATAAAATTTCAATGAGCTCCACTGGC 360
Qy 370 acacgagtgctttaaagagacccctcctacgctgggagctcgcggtaacagctgcaccc 429
Db 361 ACACGAGTGCTTTTAACAGGACCCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCC 420
Qy 430 ggagagtaatgatacaccagctttaaagtagagctcctggtcaccctttcttcttcttctt 489
Db 421 GSAGAGTAATGATCACCAGCTTTAATGTAGGAGTCTTGCTACCTTTTCTCTGCTCCCGA 480
Qy 490 tggctctgactcgtctattgccactatcttcagcagtggtgagaaacacagcgtcttct 549
Db 481 TCGGTCTGATCCTGCTCATTCGCCACTATCTTTCAGAGTGGTGAACAGACAGCTCTTCT 540
Qy 550 ctgtatctcgcgcttgagctccctcctgctgagctggaattcta 593
Db 541 CTGTATCTCGCCCGTTGGAGTCTCCTGTGCAGCTGGAAATTTCTA 584

NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. .946
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR17E23"
/note="end : T7"

BASE COUNT 200 a 169 c 226 g 302 t 49 others
ORIGIN

Query Match 2.3%; Score 45.8; DB 122; Length 946;
Best Local Similarity 61.1%; Pred. No. 0.037;
Matches 69; Conservative 3; Mismatches 41; Indels 0; Gaps 0;

Qy 1858 caagattcgatgcttttgcctagtccttagcttaactgctgattatttctgctacg 1917

Db 329 CAACTGTTCAAAAATCCTGTTATAGTTAAAGTTTAAACCTTTTGTAGTTCTTATA 270

Qy 1918 gtgtcagctcaaaaataaataaataaagctaaataaataaataaataaataa 1970

Db 269 CAGAAGATTCAATAAAGCAAGTAAGTAATATAAATAAATAAATAAATAA 217

RESULT 10

C24663 455 bp mRNA EST 28-APR-1999
LOCUS
DEFINITION C24663 Dictyostelium discoideum SS (H.Brushihara) Dictyostelium
discoideum cDNA clone SL-X066, mRNA sequence.

ACCSSION C24663

VERSION C24663.1 GI:2243084

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE Dictyostelium discoidei.

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,

Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development

JOURNAL DNA Res. 5 (6), 335-340 (1998)

MEDLINE 99156227

COMMENT On Jul 9, 1999 this sequence version replaced gi:5435125.

Contact: Hideo Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402hu@sakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'

POLYA-No.

Location/Qualifiers

1. .455

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SL-X066"

/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev_stage="slug"

BASE COUNT 262 a 56 c 20 g 117 t

ORIGIN

Query Match 2.3%; Score 45.4; DB 80; Length 455;

Best Local Similarity 56.3%; Pred. No. 0.038;

Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1820 tcagaaacacacacacacacagcgaattgtagcaccctcaagattcgcatagtctttgt 1879

Db 265 TCACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 324

Qy 1880 catagtccttagcttcttaactgctgattattatttcgtacgggtctcgcagctcaaaaataa 1939

Db 325 TTTTTCCTTGAACCCCTTAAAAAAAATTTAAATTTTAAAAACCCCAAAAAA 384

Qy 1940 tcaaatgaagctaaaaataaataaataaataaataa 1970

Db 385 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 415

RESULT 11

CNS00DKY/c

LOCUS

DEFINITION CNS00DKY 928 bp DNA GSS 04-JUN-1999

Drosophila melanogaster genome survey sequence T7 end of BAC #

BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCSSION AL071865

VERSION AL071865.1 GI:4948170

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 928)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazuhiro Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

source

1. .928

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR27A24"

/note="end : T7"

BASE COUNT 262 a 70 c 84 g 321 t 191 others

ORIGIN

Query Match

Best Local Similarity 28.9%; Pred. No. 0.054;

Matches 41; Conservative 55; Mismatches 46; Indels 0; Gaps 0;

Qy 1829 acacaaacacagcgaattgacacacctcaagattcgatgctttttgtcatgctct 1888

Db 853 AAAAAAAMCMVMMMMMMMMMMMMMTTMMHHHHHHHHHTTTTWTWTWT 794

Qy 1889 tagcttaactgctgattattttgtacggttcgagctcaaaaataaataaataa 1948

Db 793 TTTTWWWWWWWWTTTWWTWWTWWWWWWWWWWTNNNNNMCHWAWATWWWWATATTW 734

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@atson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3322 FAX: (888) 619-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com

esystems.com or
www.genomesystems.com

Location/Qualifiers
1. .202

```

1. .202
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl009-3766"
/clone_lib="Gm-cl009"

```

/lab_host="XLU10-Gold"
/note="vector: pBluescript II XR; Site1: EcoRI; Site2: XhoI; The mRNA was isolated from entire roots of 2-month-old 'Williams' plants that were greenhouse grown in 5-gallon pots. To suppress nodulation, Black Gold All-purpose potting soil was supplemented with: 0.36g/L available phosphoric acid (P2O5), 20mg/L urea N, 0.16g/L S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn, 0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following nutrients in a slow-release form (Osmocote): 0.165g/L ammonia N, 0.185g/L nitrate N, 0.35g/L available

ammonia N, 0.185g/L nitrate N, 0.35g/L available phosphoric acid, and 0.35g/L soluble potash. No nodules were visible on the roots at harvest. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dcrp, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used.

Second-strand synthesis is, the cDNA ends were polished with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 400bp cutoff, using a sizeSep 400 spin column from Pharmacia. The column eluent was then

ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with

2.28: Score 43.8: DB 70: Length 202:

Identity 64.1%; Pred. No. 0.087;
 conservative 0; Mismatches 37; Indels 0; Gaps 0

ttcataatccttaagtcttaaacctgattattttctacaggtttcgaac 1927

[illegible]

```

September  4, 2000, 00:24:48
AAAAAAAAAAAAAAAAAAAAAAAAAAAA 175

```

THIS PAGE BLANK (USPTO)